

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 19:57:16 ; Search time 79.2548 Seconds  
(without alignments)  
3347.654 Million cell updates/sec

Title: US-10-658-782-2

Perfect score: 3619

Sequence: 1 MAPITAYAQTRGLGCIIT.....PAIPDREVLRYDFDEMEEC 686

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3619	100.0	686	5	Aau76377 Hepatitis
2	3619	100.0	686	5	Aae18689 HCV-1 NS3
3	3619	100.0	686	6	Abg72261 HCV-1 NS3
4	3619	100.0	686	7	Adc06767 HCV mutan
5	3619	100.0	686	8	Adl16805 HCV NS3/4
6	3602	99.5	2261	1	Aap90164 Peptide e
7	3602	99.5	2436	1	Aap92050 HCV prote
8	3602	99.5	2436	1	Aap90288 Peptide e
9	3602	99.5	2772	3	Aab18540 Protein e
10	3602	99.5	2772	8	Adn35976 HCV CDNA
11	3602	99.5	2955	2	Aay14975 Amino aci
12	3602	99.5	2955	3	Aab18541 Polypepte
13	3602	99.5	2955	8	Adh35978 HCV CDNA
14	3602	99.5	3011	2	Aar90931 Hepatitis
15	3602	99.5	3011	2	Aaw34480 HCV polyp
16	3602	99.5	3011	2	Aaw40038 HCV polyp
17	3602	99.5	3011	2	Aae22049 Hepatitis
18	3602	99.5	3011	8	Adl23107 Hepatitis
19	3602	99.5	3011	8	Adr29357 Hepatitis
20	3600	99.5	728	5	Aae18688 NS3/4a mu
21	3600	99.5	728	7	Adc06766 HCV mutan
22	3599	99.4	2301	1	Aap92047 HCV prote
23	3595	99.3	2772	2	Aar08123 Hepatitis
24	3594	99.3	686	4	Aab62633 HCV NS3/4
25	3593	99.3	3011	2	Aar21519 Compiled

26	3590	99.2	2435	2	AAR25135	Aar25135 HCV polyp
27	3589	99.2	3011	2	AAR31621	Aar31621 Hepatitis
28	3587	99.1	3011	5	AAU84597	Aau84597 HCV polyp
29	3586	99.1	2816	2	AAR34009	Aar34009 HCV-1 pol
30	3583	99.0	1786	1	AAP90158	Aap90158 Protein 8
31	3583	99.0	2436	2	AAR28582	Aar28582 HCV amino
32	3583	99.0	2894	2	AAR70230	Aar70230 Composit
33	3580	98.9	2894	2	AAR24440	Aar24440 Composit
34	3579	98.9	1766	1	AAP92041	Aap92041 Hepatitis
35	3565	98.5	686	5	AAE21837	Aae21837 Hepatitis
36	3565	98.5	686	5	AAE19900	Aae19900 Hepatitis
37	3565	98.5	686	7	ABW00351	Abw00351 Hepatitis
38	3565	98.5	686	8	ADG47659	Adg47659 HCV NS3/4
39	3565	98.5	686	8	ADG47693	Adg47693 HCV NS3/4
40	3563	98.5	1986	8	ADR38451	Adr38451 Hepatitis
41	3561	98.4	686	5	AAE21838	Aae21838 Hepatitis
42	3561	98.4	686	5	AAE19907	Aae19907 Hepatitis
43	3561	98.4	686	7	ABW00358	Abw00358 Hepatitis
44	3561	98.4	686	8	ADG47660	Adg47660 HCV NS3/4
45	3561	98.4	3011	2	AAR40120	Aar40120 HCV genom

#### ALIGNMENTS

#### RESULT 1

AAU76377

ID AAU76377 standard; protein; 686 AA.

AC AAU76377;

DT 08-MAY-2002 (first entry)

DE Hepatitis C virus NS3/4a conformational epitope protein sequence.

KW Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
KW immunoassay solid support; multiple epitope fusion antigen; MEPA;  
KW non-structural protein; mutant; mutein.

OS Hepatitis C virus.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 403 /note= "Wild-type Thr substituted by Pro"

FT Misc-difference 404 /note= "Wild-type Ser substituted by Ile"

XX WO200196870-A2.

XX 20-DEC-2001.

XX 14-JUN-2001; 2001WO-US019156.

XX 15-JUN-2000; 2000US-0212082P.

XX 02-APR-2001; 2001US-0280811P.

XX 02-APR-2001; 2001US-0280867P.

XX (CHIR ) CHIRON CORP.

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;

XX Medina-Selby A;

XX WPI; 2002-090228/12.

XX N-PSDB; ABK15344.

XX Immunoassay solid support, useful for detecting hepatitis C virus  
XX infection in biological sample, comprises HCV NS3/4a conformational  
XX epitope and multiple epitope fusion antigen bound to the support.

XX Claim 5; Fig 3; 92pp; English.

XX The present invention relates to a new immunoassay solid support

CC

consisting essentially of at least one hepatitis C virus (HCV) NS3/4a conformational epitope and a multiple epitope fusion antigen (MEFA), bound to the support. The NS3/4a conformational epitope and/or MEFA, reacts specifically with anti-HCV antibodies present in a biological sample from an HCV-infected individual. The immunoassay of the invention is useful for detecting hepatitis C virus infection in a biological sample. The method of the invention provides a sensitive, accurate diagnostic and prognostic tool to provide adequate patient care and to prevent transmission of HCV by blood and by blood products, or by personal contact. Use of NS3/4a conformational epitope in combination with MEFA, provides a sensitive and reliable method for detecting early HCV seroconversion. Use of MEFA has the added advantages of decreasing masking problems, improving sensitivity in detecting antibodies by allowing a greater number of epitopes on a unit surface area of substrate, and improving substrate. Detection accuracy is increased and the incidence of false results is reduced because of the identification and the use of highly immunogenic HCV antigens which are present during the early stages of HCV seroconversion. The present amino acid sequence represents the non-structural protein NS3/4a conformational epitope of the invention.

XX  
SQ Sequence 686 AA;

Query Match	100.0%	Score 3619;	DB 5;	Length 686;
Best Local Similarity	100.0%;	Pred. No. 2.5e-306;		
Matches 686;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MAPITTAQAQTRGLGCGIITSLTRGDKNQVEGEVQIVSTAAQTFLATCINGVCWTVTHGA	60	
Db	1	MAPITTAQAQTRGLGCGIITSLTRGDKNQVEGEVQIVSTAAQTFLATCINGVCWTVTHGA	60	
Qy	61	GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSLSLTCGSSDLVLTVRHADVIPVRRR	120	
Db	61	GTRTIASPKGPVIOMYTNVDQDLVGPAPQGSRSLSLTCGSSDLVLTVRHADVIPVRRR	120	
Qy	121	GDGRGSLLSRPISLYLKSSGGPPLLCPAGHAVGIFRAAVCTRGVAKAVDPFIPVENLETTM	180	
Db	121	GDGRGSLLSRPISLYLKSSGGPPLLCPAGHAVGIFRAAVCTRGVAKAVDPFIPVENLETTM	180	
Qy	181	RSPVFTDNSPPVVPQSFQVAHLHAPTGSKGSKTKVPAAYAAQGYKVLVLPNSVAATLQFG	240	
Db	181	RSPVFTDNSPPVVPQSFQVAHLHAPTGSKGSKTKVPAAYAAQGYKVLVLPNSVAATLQFG	240	
Qy	241	AYMSKAHGIDPNIRKTGVRTITGTSPIYTVSYTKFLADGGCGGGAYDIIICDECHSTDATS	300	
Db	241	AYMSKAHGIDPNIRKTGVRTITGTSPIYTVSYTKFLADGGCGGGAYDIIICDECHSTDATS	300	
Qy	301	ILGTGTVLDOAETAGARLVLATATPGSVTVPHPNIEEVALSTTGELPFYGKAIPLEVI	360	
Db	301	ILGTGTVLDOAETAGARLVLATATPGSVTVPHPNIEEVALSTTGELPFYGKAIPLEVI	360	
Qy	361	KGRHLLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYT	420	
Db	361	KGRHLLIFCHSKKKCDELAALKVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYT	420	
Qy	421	GDSDSVLDNCNCTVQTQTVDFSLDPTFTTETITLPQDAVSRQRRGRTGRGPGIYRFVAPG	480	
Db	421	GDSDSVLDNCNCTVQTQTVDFSLDPTFTTETITLPQDAVSRQRRGRTGRGPGIYRFVAPG	480	
Qy	481	ERSGMFDSVSLCBYDAGCAWYBELTPAETTVRLRAYMNTPLGPVCDHLEFWEGVFTGL	540	
Db	481	ERSGMFDSVSLCBYDAGCAWYBELTPAETTVRLRAYMNTPLGPVCDHLEFWEGVFTGL	540	
Qy	541	THIDAHFLSQTKSGENLPVLVYAQTVCARAAQPPPSWDQMKCLIRLKPRTLHGPTPLL	600	
Db	541	THIDAHFLSQTKSGENLPVLVYAQTVCARAAQPPPSWDQMKCLIRLKPRTLHGPTPLL	600	
Qy	601	YRLGAVONEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCVCWIVGR	660	
Db	601	YRLGAVONEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCVCWIVGR	660	
Qy	661	VVLSGKPAIIIPDREVLRYEFDEMEEC	686	



Qy 481 ERPSGMFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGLPVCDHLEFEGVFTGL 540  
Db 481 ERPSGMFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGLPVCDHLEFEGVFTGL 540  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLPTLHGPTPL 600  
Db 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLPTLHGPTPL 600  
Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCISGTCVWVGR 660  
Db 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCISGTCVWVGR 660  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 661 VVLSGKPAIIPDREVLYREFDEMEEC 686

RESULT 4  
ADC06767  
ID ADC06767 standard; protein; 686 AA.  
XX AC  
XX AD06767;  
XX 18-DEC-2003 (first entry)  
XX HCV mutant conformational NS3/4a epitope protein T403P/S404I.  
XX immunoassay solid support; HCV; NS3/4a; non-structural;  
KW non-A, non-B hepatitis; NANB; conformational epitope; mutant; mutein.  
XX OS  
XX Synthetic.  
OS Hepatitis C virus.  
XX PH  
XX Key Location/Qualifiers  
FT Misc-difference 303 /note= "Wild-type Thr replaced by Pro"  
FT FT  
FT Misc-difference 304 /note= "Wild-type Ser replaced by Ile"  
XX  
XX US2002192639-A1.  
XX  
XX 19-DEC-2002.  
XX  
XX 14-JUN-2001; 2001US-00881239.  
XX  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX  
XX (CHIE/) CHIEN D Y.  
PA (ARCA/) ARCANGEL P.  
PA (TAND/) TANDESKE L.  
PA (GEOR/) GEORGE-NASCIMENTO C.  
PA (COIT/) COIT D.  
PA (MEDI/) MEDINA-SELBY A.  
XX  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX  
XX WPI; 2003-644609/61.  
DR N-PSDB; ADC06768.  
XX  
XX Immunoassay solid support for detecting hepatitis C virus infection in  
PT biological samples, comprises a hepatitis C virus anti-core antibody and  
PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
PT antibody.  
XX  
XX Claim 6; Fig 4; 40pp; English.  
XX  
XX The invention relates to a novel immunoassay solid support comprising at  
CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.

CC The system of the invention may be useful for detecting HCV infection in  
CC a biological sample and for treating or detecting non-A, non-B hepatitis  
CC (NANB hepatitis). The current sequence is that of the HCV mutant  
CC conformational NS3/4a epitope protein of the invention which contains  
CC T403P/S404I mutations.  
XX  
SQ Sequence 686 AA;  
Query Match 100.0%; Score 3619; DB 7; Length 686;  
Best Local Similarity 100.0%; Pred. No. 2.5e-306;  
Matches 686; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Db 1 MAPITAYAOQTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Qy 61 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTVRHADVIPVRR 120  
Db 61 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTVRHADVIPVRR 120  
Qy 121 GDSRGLSPRPISYILKSGSGGPLLCPAGHANGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 121 GDSRGLSPRPISYILKSGSGGPLLCPAGHANGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Qy 181 RSPVFTDNSSPPVVPQSFQVAHLHPTGSGKSTKVPAAVAAQYKVLNPSVAATLFGF 240  
Db 181 RSPVFTDNSSPPVVPQSFQVAHLHPTGSGKSTKVPAAVAAQYKVLNPSVAATLFGF 240  
Qy 241 AYMSKAHGIDPNRTGVRTITTSPTITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Db 241 AYMSKAHGIDPNRTGVRTITTSPTITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEIPYKAIPLV 360  
Db 301 ILGIGTVLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEIPYKAIPLV 360  
Qy 361 KGRHLIFCHSKKCKDELAALVALGINAVAYYRGDVSVIPIGDVVVVATDALMTGYT 420  
Db 361 KGRHLIFCHSKKCKDELAALVALGINAVAYYRGDVSVIPIGDVVVVATDALMTGYT 420  
Qy 421 GDFDSVIDCNTCTVTQVDFSLDPTFTTITLTPQDAVSRTRGRGKPGIYRFVAPG 480  
Db 421 GDFDSVIDCNTCTVTQVDFSLDPTFTTITLTPQDAVSRTRGRGKPGIYRFVAPG 480  
Qy 481 ERPSGMFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGLPVCDHLEFEGVFTGL 540  
Db 481 ERPSGMFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGLPVCDHLEFEGVFTGL 540  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLPTLHGPTPL 600  
Db 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLPTLHGPTPL 600  
Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCISGTCVWVGR 660  
Db 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCISGTCVWVGR 660  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 661 VVLSGKPAIIPDREVLYREFDEMEEC 686

RESULT 5  
ADL66805  
ID ADL66805 standard; protein; 686 AA.  
XX AC  
XX ADL66805;  
XX 03-JUN-2004 (first entry)  
XX  
XX HCV NS3/4a conformational epitope.  
XX  
XX HCV; NS3/4a conformational epitope; HCV antigen; HCV polyprotein;  
KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection.



XX OS Hepatitis C virus.  
 XX PN WO2004021871-A2.  
 XX PD 18-MAR-2004.  
 XX PF 08-SEP-2003; 2003WO-US028071.  
 XX PR 09-SEP-2002; 2002US-0409515P.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PI Arcangel P, Chien D;  
 XX DR WPI; 2004-248333/23.  
 XX DR N-PSDB; ADL66804.  
 XX PT Detecting hepatitis C virus (HCV) infection in a biological sample by  
 PT detecting complexes formed between the HCV antibody and the antigens from  
 PT the first region of the HCV polyprotein and the multiple epitope fusion  
 PT antigen (MEFA).  
 XX PS Claim 5; SEQ ID NO 2; 93pp; English.  
 XX  
 CC The invention relates to a method of detecting hepatitis C virus (HCV)  
 CC infection in a biological sample. The method comprises providing an  
 CC immunocassay solid support comprising HCV antigens bound to it, where the  
 CC HCV antigens comprise one or more isolated antigens form a first region  
 CC of the HCV polyprotein, combining a biological sample with the solid  
 CC support under conditions that allow HCV antibodies, when present in the  
 CC biological sample, to bind to the one or more HCV antigens, adding to the  
 CC solid support a detectably labelled HCV multiple epitope fusion antigen  
 CC (MEFA), where the labelled MEFA comprises at least one epitope from the  
 CC same region of the HCV polyprotein as the one or more isolated antigens,  
 CC where the MEFA binds to the bound HCV antibody, and detecting complexes  
 CC formed between the HCV antibody and the one or more antigens from the  
 CC first region of the HCV polyprotein and the MEFA, if any, as an  
 CC indication of HCV infection in the biological sample. The method is  
 CC useful for detecting hepatitis C virus (HCV) infection in a biological  
 CC sample. This sequence represents the NS3/4a conformational epitope used  
 CC in the scope of the invention.  
 XX SQ Sequence 686 AA;

Query Match 100.0%; Score 3619; DB 8; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-306; Mismatches 0; Indels 0; Gaps 0;  
 Matches 686; Conservative 0;

QY 1 MAPITAYAAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60  
 DB 1 MAPITAYAAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCTVYHGA 60

QY 61 GTRTIASPKGVIOYNTNVDDQLGVGPAQPSRSLTPTCTGSSDLVLTTRHADVIVRRR 120  
 DB 61 GTRTIASPKGVIOYNTNVDDQLGVGPAQPSRSLTPTCTGSSDLVLTTRHADVIVRRR 120

QY 121 GDSRGLSPRPISVLYKSSGGPLLCFAGHAGVIFRAACVTRGVAKAVDFIPVENLETTM 180  
 DB 121 GDSRGLSPRPISVLYKSSGGPLLCFAGHAGVIFRAACVTRGVAKAVDFIPVENLETTM 180

QY 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240  
 DB 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240

QY 241 AYMSKAHGIDNIRTVGRTITITGSPITVSTYTKFLADGGCGGGAYDIIICDECHSTDATS 300  
 DB 241 AYMSKAHGIDNIRTVGRTITITGSPITVSTYTKFLADGGCGGGAYDIIICDECHSTDATS 300

QY 301 ILGIGTVLDQAEATAGARLVLAATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVI 360  
 DB 301 ILGIGTVLDQAEATAGARLVLAATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVI 360

QY 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGDLDSVIPPIDGVVVVATDALMTGYT 420  
 DB 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGDLDSVIPPIDGVVVVATDALMTGYT 420

QY 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRRTORRGRTGRGKPGIYRFVAPG 480  
 DB 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRRTORRGRTGRGKPGIYRFVAPG 480

QY 481 ERPSGMFSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTTGLPVCQDHLFEFWEVFTGL 540  
 DB 481 ERPSGMFSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTTGLPVCQDHLFEFWEVFTGL 540

QY 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQPPSWDMMKCLIRLKPTELHGPTPL 600  
 DB 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQPPSWDMMKCLIRLKPTELHGPTPL 600

QY 601 YRLGAVONEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIGR 660  
 DB 601 YRLGAVONEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIGR 660

QY 661 VLSGKPAIIPDREVLYREFDEMEEC 686  
 DB 661 VLSGKPAIIPDREVLYREFDEMEEC 686

RESULT 6  
 AAP90164  
 ID AAP90164 standard; protein; 2261 AA.  
 XX AC AAP90164;  
 XX DT 25-MAR-2003 (revised)  
 DT 01-NOV-1989 (first entry)  
 XX DE Peptide encoded by composite hepatitis C virus cDNA.  
 XX KW Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.  
 XX OS Pan troglodytes.  
 XX PN GB2212511-A.  
 XX PD 26-JUL-1989.  
 XX PF 18-NOV-1988; 88GB-00027024.  
 XX PR 18-NOV-1987; 87US-00122714.  
 PR 30-DEC-1987; 87US-00139886.  
 PR 26-FEB-1988; 88US-00161072.  
 PR 26-OCT-1988; 88US-00263584.  
 XX PA (CHIR ) CHIRON CORP.  
 XX PI Houghton M, Choo QL, Kuo G;  
 XX WPI; 1989-215054/30.  
 DR N-PSDB; AAN90331.  
 XX PT Hepatitis C virus gene - used for prodn. of polynucleotide probes  
 PT polypeptide(s) and antibodies for diagnosis, prevention and treatment of  
 PT infection.  
 XX PS Disclosure; Fig 32; 30pp; English.  
 CC The sequence is the peptide encoded by the composite hepatitis C virus  
 CC (HCV) cDNA of AAN90331. The polypeptides are used to diagnose HCV-induced  
 CC NANBH, to raise antibodies for immunoassay or treatment, or to produce  
 CC vaccines. (Updated on 25-MAR-2003 to correct PR field.)  
 XX SQ Sequence 2261 AA;

Query Match 99.5%; Score 3602; DB 1; Length 2261;  
 Best Local Similarity 99.6%; Pred. No. 4.4e-304;

Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAYAAQTGRLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db	401 LAPITAYAAQTGRLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 460
Qy	61 GTRTIIASPKGPVIMQNTYNDQDLVGNPAPQSGRSLSLTPCTCGSSDLYLVTRHADVIPIVRRR 120
Db	461 GTRTIIASPKGPVIMQNTYNDQDLVGNPAPQSGRSLSLTPCTCGSSDLYLVTRHADVIPIVRRR 520
Qy	121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db	521 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 580
Qy	181 RSPVFTDNSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNPSVAATLGF 240
Db	581 RSPVFTDNSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNPSVAATLGF 640
Qy	241 AYMSKAHGIDPNIRTCVRIITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300
Db	641 AYMSKAHGIDPNIRTCVRIITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 700
Qy	301 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db	701 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 760
Qy	361 KGRHLIFCHSKKKCDLAALVALGINAVAYYRGDLVSVIPIGDVWVATDALMTGYT 420
Db	761 KGRHLIFCHSKKKCDLAALVALGINAVAYYRGDLVSVIPIGDVWVATDALMTGYT 820
Qy	421 GDFDSVIDCNTCTVQTVDPSLDPTFTTETITLPODAVSRQRRGTRGKPGIYRFVAPG 480
Db	821 GDFDSVIDCNTCTVQTVDPSLDPTFTTETITLPODAVSRQRRGTRGKPGIYRFVAPG 880
Qy	481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 540
Db	881 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 940
Qy	541 THIDAHFLSQTOSGENLPYLVAIQATVCARAQPPSWDMWKCLIRLKLHGTPL 600
Db	941 THIDAHFLSQTOSGENLPYLVAIQATVCARAQPPSWDMWKCLIRLKLHGTPL 1000
Qy	601 YRLGAVQNEITLTHPVTKYIMTCKMSADLEVTSTWLVGGVLAALAAAYCISLSTGCWTVGR 660
Db	1001 YRLGAVQNEITLTHPVTKYIMTCKMSADLEVTSTWLVGGVLAALAAAYCISLSTGCWTVGR 1060
Qy	661 VVLSGKPAIIPDREVLRYFDEMEEC 686
Db	1061 VVLSGKPAIIPDREVLRYFDEMEEC 1086
RESULT 7	
AAP92050	
ID	AAP92050 standard; protein; 2436 AA.
XX	
AC	AAP92050;
XX	
DT	09-SEP-2004 (revised)
DT	25-MAR-2003 (revised)
DT	02-MAR-1990 (first entry)
XX	
DE	HCV protein of the cDNA inserts in clones K9-1 through 15e.
XX	
KW	Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).
XX	
OS	Hepatitis C virus.
OS	Unidentified.
XX	
PN	EP318216-A.
XX	
PD	31-MAY-1989.
XX	
PF	18-NOV-1988; 88EP-00310922.

XX	18-NOV-1987; 87US-00122714.
PR	30-DEC-1987; 87US-00139886.
PR	26-FEB-1988; 88US-00161072.
PR	06-MAY-1988; 88US-00131263.
PR	26-OCT-1988; 88US-00263584.
PR	14-NOV-1988; 88US-00271450.
XX	
PA	(CHIR ) CHIRON CORP.
PA	(CHIR ) CHIRON CORP.
XX	
PI	Houghton M, Choo QL, Kuo G;
XX	
DR	WPI; 1989-159274/22.
DR	N-PSDB; AAN92106.
XX	
PT	Purified hepatitis C virus - and associated nucleic acids and
PT	polypeptide(s).
XX	
PS	Claim 13; Fig 47-1-47-8; 139pp; English.
XX	
CC	It is the sequence encoded in the open reading frame of hepatitis C virus
CC	(HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could
CC	be used in immunoassay reagents and vaccines and to generate antibodies
CC	useful in diagnosis and passive immunotherapy for HCV infection/non-A,
CC	non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC	on 25-MAR-2003 to correct PI field.)
CC	
CC	Revised record issued on 09-SEP-2004 : Correction to DE line
XX	
SQ	Sequence 2436 AA;
Query Match 99.5%; Score 3602; DB 1; Length 2436;	
Best Local Similarity 99.6%; Pred. No. 5e-304; Mismatches 0; Gaps 0;	
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAYAAQTGRLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db	576 LAPITAYAAQTGRLGCIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 635
Qy	61 GTRTIIASPKGPVIMQNTYNDQDLVGNPAPQSGRSLSLTPCTCGSSDLYLVTRHADVIPIVRRR 120
Db	636 GTRTIIASPKGPVIMQNTYNDQDLVGNPAPQSGRSLSLTPCTCGSSDLYLVTRHADVIPIVRRR 695
Qy	121 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db	696 GDSRGSLLSPRPISYILKSGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755
Qy	181 RSPVFTDNSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNPSVAATLGF 240
Db	756 RSPVFTDNSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNPSVAATLGF 815
Qy	241 AYMSKAHGIDPNIRTCVRIITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 300
Db	816 AYMSKAHGIDPNIRTCVRIITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATS 875
Qy	301 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 360
Db	876 ILGIGTVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI 935
Qy	361 KGRHLIFCHSKKKCDLAALVALGINAVAYYRGDLVSVIPIGDVWVATDALMTGYT 420
Db	936 KGRHLIFCHSKKKCDLAALVALGINAVAYYRGDLVSVIPIGDVWVATDALMTGYT 995
Qy	421 GDFDSVIDCNTCTVQTVDPSLDPTFTTETITLPODAVSRQRRGTRGKPGIYRFVAPG 480
Db	996 GDFDSVIDCNTCTVQTVDPSLDPTFTTETITLPODAVSRQRRGTRGKPGIYRFVAPG 1055
Qy	481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 540
Db	1056 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTTGPLVPCQDHLFEFEGVFTGL 1115
Qy	541 THIDAHFLSQTOSGENLPYLVAIQATVCARAQPPSWDMWKCLIRLKLHGTPL 600

```

Db      1116 THIDAHFLSQTQSGENLPLVAYQATVCARAQAPPSWDMKCLIRLKPTLHGPTPLL 1175
Qy      601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCLSTGCWIVGR 660
Db      1176 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCLSTGCWIVGR 1235
Qy      661 VVLSGKPAIIPDREVLRYREFDEMEEC 686
Db      1236 VVLSGKPAIIPDREVLRYREFDEMEEC 1261

RESULT 8
AAP90288
ID AAP90288 standard; protein; 2436 AA.
XX AC
XX AAP90288;
XX DT 25-MAR-2003 (revised)
XX DT 19-JUL-2001 (revised)
XX DT 01-NOV-1989 (first entry)
XX XX
XX Peptide encoded by composite hepatitis C cDNA.
XX Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.
XX Pan troglodytes.
XX OS
XX GB2212511-A.
XX PD
XX PF 18-NOV-1988; 88GB-00027024.
XX PR 18-NOV-1987; 87US-00122714.
XX PR 30-DEC-1987; 87US-00139886.
XX PR 26-FEB-1988; 88US-00161072.
XX PR 26-OCT-1988; 88US-00263584.
XX PA
XX (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX WPI; 1989-215054/30.
XX DR N-PSDB; AAN90336.
XX XX
XX Hepatitis C virus gene - used for prodn. of polynucleotide probes
XX polypeptide(s) and antibodies for diagnosis, prevention and treatment of
XX infection.
XX PS Disclosure; Fig 47-1 to 47-8; 30pp; English.
XX CC The sequence is the peptide encoded by the composite hepatitis C virus
XX (HCV) cDNA of AAN90336. The polypeptides are used to diagnose HCV-induced
XX NANBH, to raise antibodies for immunoassay or treatment, or to produce
XX vaccines. (N.B. This record was resubmitted to correct errors in the
XX sequence.) (Updated on 25-MAR-2003 to correct PR field.)
XX SQ Sequence 2436 AA;

Query Match 99.5%; Score 3602; DB 1; Length 2436;
Best Local Similarity 99.6%; Pred. No. 5e-304;
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MAPITAAQTRGLLGCILTSITGRDNQKQVEGEVQIVSTAQTFLATCINGCVCTVYHGA 60
Db      576 LAPITAAQTRGLLGCILTSITGRDNQKQVEGEVQIVSTAQTFLATCINGCVCTVYHGA 635
Qy      61 GTRTIASPKGPVQIOMYTNVDQDLVGPAPQGSRLTPTCTGSSDLVLRHADVIPVRRR 120
Db      636 GTRTIASPKGPVQIOMYTNVDQDLVGPAPQGSRLTPTCTGSSDLVLRHADVIPVRRR 695
Qy      121 GDSRGSLLSPRISYLGKSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180

```

```

Db      696 GDSRGSLLSPRISYLGKSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 755
Qy      181 RSPVFTDNSSPPVPOSFQVAHLHAPTSGSKSTKPAAYAAQGYKVLVLPNSVAATLGFG 240
Db      756 RSPVFTDNSSPPVPOSFQVAHLHAPTSGSKSTKPAAYAAQGYKVLVLPNSVAATLGFG 815
Qy      241 AYMSKAHGDIPNIRTGVRTITTCSPITYSTYTGFLADGGCGSGAYDIIICDECHSTDATS 300
Db      816 AYMSKAHGDIPNIRTGVRTITTCSPITYSTYTGFLADGGCGSGAYDIIICDECHSTDATS 875
Qy      301 ILGIGTVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTTGBIPYGAIPLEVI 360
Db      876 ILGIGTVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTTGBIPYGAIPLEVI 935
Qy      361 KGGRHILFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 420
Db      936 KGGRHILFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVIPPTSGDVVVVATDALMTGYT 995
Qy      421 GDFDSVIDCNTCTVTQVDFSLDPTFTIETITLPODAVSRQRRGTRGRGKPGIYRFVAPG 480
Db      996 GDFDSVIDCNTCTVTQVDFSLDPTFTIETITLPODAVSRQRRGTRGRGKPGIYRFVAPG 1055
Qy      481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCCDHLEFEGVFTGL 540
Db      1056 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCCDHLEFEGVFTGL 1115
Qy      541 THIDAHFLSQTQSGENLPLVAYQATVCARAQAPPSWDMKCLIRLKPTLHGPTPLL 600
Db      1116 THIDAHFLSQTQSGENLPLVAYQATVCARAQAPPSWDMKCLIRLKPTLHGPTPLL 1175
Qy      601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCLSTGCWIVGR 660
Db      1176 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAAAYCLSTGCWIVGR 1235
Qy      661 VVLSGKPAIIPDREVLRYREFDEMEEC 686
Db      1236 VVLSGKPAIIPDREVLRYREFDEMEEC 1261

RESULT 9
AAB18540
ID AAB18540 standard; protein; 2772 AA.
XX AC
XX AAB18540;
XX DT 15-JAN-2001 (first entry)
XX DE Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.
XX KW Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;
XX KW viral infectivity; viral replication.
XX OS Hepatitis C virus.
XX PN EPI034785-A2.
XX PD 13-SEP-2000.
XX PF 16-MAR-1990; 2000EP-00109602.
XX PR 17-MAR-1989; 89US-00325338.
XX PR 20-APR-1989; 89US-00341334.
XX PR 18-MAY-1989; 89US-00355002.
XX PR 16-MAR-1990; 90EP-00302866.
XX PA (CHIR ) CHIRON CORP.
XX PI Houghton M, Choo Q, Kuo G;
XX WPI; 2000-566891/53.
XX DR N-PSDB; AAA75296.
XX

```

PT Novel composition comprising a hepatitis C virus antisense polynucleotide  
PT which is complementary to or corresponds to a sense strand of the virus  
PT genome, and selectively hybridizes to it.

PS Example; Fig 16; 75pp; English.

XX The specification describes a pharmaceutical composition which comprises  
CC a hepatitis C virus (HCV) antisense polynucleotide. The HCV is  
CC characterized by a positive stranded RNA genome which has 40% homology at  
CC the polypeptide level to a HCV polypeptide. The antisense polynucleotide  
CC binds to cellular polynucleotides which enhance and/or are required for  
CC viral infectivity, replicative ability or chronicity. The antisense  
CC polynucleotides may also be designed to bind with high specificity, to be  
CC of increased stability, to be stable and to have low toxicity. The  
CC composition also comprises an agent which causes viral RNA to be  
CC inactive. The composition is used for preventing HCV replication in a  
CC system. The present sequence is encoded by a novel HCV cDNA sequence,  
CC which is used in the course of the invention

XX SQ Sequence 2772 AA;

Query Match 99.5%; Score 3602; DB 3; Length 2772;  
Best Local Similarity 99.6%; Pred. No. 6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
DB 912 LAPITAYAOQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 971  
QY 61 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120  
DB 972 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 1031  
QY 121 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1032 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1091  
QY 181 RSPVFTDNNSPVPVPSFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 240  
DB 1092 RSPVFTDNNSPVPVPSFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 1151  
QY 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS 300  
DB 1152 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS 1211  
QY 301 ILGIGTVLDAQETAGARLVLTATPPGSGTVPHPNIEEVALSTTGEIPFYGKAIPLEVI 360  
DB 1212 ILGIGTVLDAQETAGARLVLTATPPGSGTVPHPNIEEVALSTTGEIPFYGKAIPLEVI 1271  
QY 361 KGGRHLLIFCHSKKCDLAALVALGINAVAYYRGLOVSVIPPIGDVVVATDALMTGYT 420  
DB 1272 KGGRHLLIFCHSKKCDLAALVALGINAVAYYRGLOVSVIPPIGDVVVATDALMTGYT 1331  
QY 421 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRQRRORTGRGKPGIYRFVAPG 480  
DB 1332 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRQRRORTGRGKPGIYRFVAPG 1391  
QY 481 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMYNTPLPVCODHLEFEGVFTGL 540  
DB 1392 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMYNTPLPVCODHLEFEGVFTGL 1451  
QY 541 THIDAHFLSOTKQSGENLPYLVAQVCAQAAPPSPSDQMWKCLIRLKLPTLHGPTPL 600  
DB 1452 THIDAHFLSOTKQSGENLPYLVAQVCAQAAPPSPSDQMWKCLIRLKLPTLHGPTPL 1511  
QY 601 YRLGAVQNEITLHPVTKYIMTCMSADLEVVSTWLVGGVLAALAYCLSTGCWVIVGR 660  
DB 1512 YRLGAVQNEITLHPVTKYIMTCMSADLEVVSTWLVGGVLAALAYCLSTGCWVIVGR 1571  
QY 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
DB 1572 VVLSGKPAIIPDREVLYREFDEMEEC 1597

RESULT 10  
ADN35976

ID ADN35976 standard; protein; 2772 AA.

XX AC ADN35976;

XX DT 17-JUN-2004 (first entry)

XX DE HCV cDNA clone #1 protein.

XX KW Antiviral; Vaccine; hepatitis C virus infection; HCV infection.

XX OS Hepatitis C virus.

XX PN EP1394255-A2.

XX PD 03-MAR-2004.

XX PF 16-MAR-1990; 2003EP-00016585.

XX PR 17-MAR-1989; 89US-00325338.

XX PR 20-APR-1989; 89US-00341334.

XX PR 18-MAY-1989; 89US-00355002.

XX PR 16-MAR-1990; 90EP-00302866.

XX PA (CHIR ) CHIRON CORP.

XX PI Houghton M, Choo Q, Kuo G;

XX DR WPI; 2004-193149/19.

XX DR N-PSDB; ADN35977.

XX PT Novel purified hepatitis C virus polypeptide comprising epitope encoded  
PT by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C  
PT virus.

XX PS Example 1; Fig 16; 79pp; English.

XX CC The present invention relates to hepatitis C virus (HCV) proteins and  
CC cDNA sequences. The sequences are useful in immunoassays for detecting  
CC antibodies directed against HCV antigen; preparing host cells transformed  
CC with a recombinant polynucleotide; screening antiviral agents and  
CC determining the effect of antiviral agent in inhibiting viral replication  
CC in cell culture system; and developing vaccine for treating HCV  
CC infection.

XX SQ Sequence 2772 AA;

Query Match 99.5%; Score 3602; DB 8; Length 2772;  
Best Local Similarity 99.6%; Pred. No. 6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
DB 912 LAPITAYAOQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 971  
QY 61 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 120  
DB 972 GTRTIASPKGPVIQMTYTNVDQDLVGPAPQGSRLTPTCTCGSSDLVLTTRHADVIPVRRR 1031  
QY 121 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1032 GDSRGSLLSPRISYILKSGSGGGLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1091  
QY 181 RSPVFTDNNSPVPVPSFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 240  
DB 1092 RSPVFTDNNSPVPVPSFQVAHLHAPTGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 1151  
QY 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS 300  
DB 1152 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS 1211

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 360  
Db 1212 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 1271  
Qy 361 KGGHLLIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPPTIGDVVVATDALMTGYT 420  
Db 1272 KGGHLLIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPPTIGDVVVATDALMTGYT 1331  
Qy 421 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 480  
Db 1332 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 1391  
Qy 481 ERPSGMFDSVLCYDAGCAWYELTTPAETVRLRAYMNTFGLPVCODHLEFEGVFTGL 540  
Db 1392 ERPSGMFDSVLCYDAGCAWYELTTPAETVRLRAYMNTFGLPVCODHLEFEGVFTGL 1451  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTRLHGPTPL 600  
Db 1452 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTRLHGPTPL 1511  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWIVGR 660  
Db 1512 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWIVGR 1571  
Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686  
Db 1572 VVLSGKPAIIPDREVLYREFDEMEEC 1597

## RESULT 11

AA114975  
ID AA114975 standard; protein; 2955 AA.

XX AC AA114975;

DT 20-MAR-2003 (revised)  
DT 08-NOV-1999 (first entry)

XX DE Amino acid sequence of HCV-1 ORF.

XX KW Hepatitis C virus; HCV; J1; J7; HCV-1; non-A, non-B HCV; NANBH;  
XX KW HCV infection; vaccine.

XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers

FT Misc-difference 441

FT Misc-difference 461

FT Misc-difference 461

XX EP939128-A2.

XX PD 01-SEP-1999.

XX PF 17-SEP-1990; 99EP-00101746.

XX PR 15-SEP-1989; 89US-00408045.

XX PR 21-DEC-1989; 89US-00456142.

XX PR 17-SEP-1990; 90EP-00310149.

XX PA (OYAA/) OYA A.

XX PA (CHIR ) CHIRON CORP.

XX PI Miyamura T, Saito I, Houghton M, Weiner AJ, Han J, Kolberg JA;  
XX PI Cha T, Irvine BD;

XX DR WPI; 1999-480843/41.

XX DR N-PSDB; AA207656.

XX PT New Hepatitis C Virus isolates, useful for diagnosis of hepatitis  
XX PT infections and development of vaccines.

PS Disclosure; Fig 12; 132pp; English.

XX CC The invention provides two new isolates of hepatitis C virus (HCV), J1  
CC and J7. These two isolates comprise nucleotide and amino acid sequences  
CC that are distinct from the HCV isolate HCV-1. The nucleotide sequences  
CC may be used to detect non-A, non-B HCV (NANBH) polynucleotides by  
CC hybridisation for diagnosis of NANBH infections. They may also be used to  
CC screen blood donors, donated blood and blood products for this infection.  
CC The isolates may also be used to isolate other naturally occurring  
CC variants of the virus. The polypeptides may be used as a vaccine for  
CC administration to patients to protect against infection with NANBH. The  
CC present sequence represents the amino acid sequence of HCV-1 ORF.  
CC (Updated on 20-MAR-2003 to correct PF field.) (Updated on 20-MAR-2003 to  
CC correct PR field.)  
XX CC

SQ Sequence 2955 AA;

Query Match 99.5%; Score 3602; DB 2; Length 2955;

Best Local Similarity 99.6%; Pred. No. 6.6e-304;

Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60

Db 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085

Qy 61 GTRTIASPKGPVIQMTNVDDQDLVGHAPAGQSSSLTPCTCGSSDLYLVTRHADVIPVRRR 120

Db 1086 GTRTIASPKGPVIQMTNVDDQDLVGHAPAGQSSSLTPCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRISYLLKSGSGGGLPCAGHAGVIFRAAVCTRGVAKAVDIPVENLETTM 180

Db 1146 GDSRGSLLSPRISYLLKSGSGGGLPCAGHAGVIFRAAVCTRGVAKAVDIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240

Db 1206 RSPVFTDNSSPPVQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265

Qy 241 AYMSKAHGIDPNIRTVRTITTTGSPITYSYGKFLADGGCGGAYDIIICDECHSTDATS 300

Db 1266 AYMSKAHGIDPNIRTVRTITTTGSPITYSYGKFLADGGCGGAYDIIICDECHSTDATS 1325

Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 360

Db 1326 ILGIGTVLDOAETAGARLVVLTATPPGCVTVPHNPNEEVALSTTGBIPFYGKAIPLEVI 1385

Qy 361 KGGHLLIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPPTIGDVVVATDALMTGYT 420

Db 1386 KGGHLLIFCHSKKCDLAAKLVALGINAVAYRGLDVSVIPPTIGDVVVATDALMTGYT 1445

Qy 421 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 480

Db 1446 GDFSDVIDCNTCTVTQTVDVDFSLDPTFTTITLPODAVSRRTQRRGRTGKPGIYRFVAPG 1505

Qy 481 ERPSGMFDSVLCYDAGCAWYELTTPAETVRLRAYMNTFGLPVCODHLEFEGVFTGL 540

Db 1506 ERPSGMFDSVLCYDAGCAWYELTTPAETVRLRAYMNTFGLPVCODHLEFEGVFTGL 1565

Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTRLHGPTPL 600

Db 1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTRLHGPTPL 1625

Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWIVGR 660

Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTSTWLVGGVLAALAAAYCLSTGCWIVGR 1685

Qy 661 VVLSGKPAIIPDREVLYREFDEMEEC 686

Db 1686 VVLSGKPAIIPDREVLYREFDEMEEC 1711

## RESULT 12

AA18541

ID AA18541 standard; protein; 2955 AA.

DB	1266	AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGFADCGCGGGAYDIIICDECHSTDATS	1325
QY	301	ILIGITVLDQAEATAGARLVWLATATPPGSGVTVPVPHENIEEVALSTTGEIPFYGKAIPLEVI	360
DB	1326	ILIGITVLDQAEATAGARLVWLATATPPGSGVTVPVPHENIEEVALSTTGEIPFYGKAIPLEVI	1385
QY	361	KGGRHLIFCHSKKKCDELAALVALGINAVAYRGDVSIVPPIGDVVVVATDALMTGYT	420
DB	1386	KGGRHLIFCHSKKKCDELAALVALGINAVAYRGDVSIVPPIGDVVVVATDALMTGYT	1445
QY	421	GDPSVIDCNTCVTQVDFSLDPTETITLPODAVSRTORRGTGRGKPGIYRFVAPG	480
DB	1446	GDPSVIDCNTCVTQVDFSLDPTETITLPODAVSRTORRGTGRGKPGIYRFVAPG	1505
QY	481	ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHFLEWEGVFTGL	540
DB	1506	ERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHFLEWEGVFTGL	1565
QY	541	THIDAFSLQTKOSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLHGPTPLL	600
DB	1566	THIDAFSLQTKOSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKLHGPTPLL	1625
QY	601	YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCVVIVGR	660
DB	1626	YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCVVIVGR	1685
QY	661	VVLGSKPAIIPDREVLRYREFDEMEEC	686
DB	1686	VVLGSKPAIIPDREVLRYREFDEMEEC	1711

RESULT 13  
ADN35978

ID	ADN35978	standard; protein; 2955 AA.
XX	ADN35978;	
AC	ADN35978;	
DT	17-JUN-2004	(first entry)
XX	HCV cDNA clone #2 protein.	
DE	Antiviral; Vaccine; hepatitis C virus infection; HCV infection.	
XX	Hepatitis C virus.	
OS	EP1394255-A2.	
PN	03-MAR-2004.	
XX	16-MAR-1990; 2003EP-00016585.	
XX	17-MAR-1989; 89US-00325338.	
PR	20-APR-1989; 89US-00341334.	
PR	18-MAY-1989; 89US-00355002.	
PR	16-MAR-1990; 90EP-00302866.	
XX	(CHIR ) CHIRON CORP.	
PA	Houghton M, Choo Q, Kuo G;	
PI	WPI; 2004-193149/19.	
XX	N-PSDB; ADN35978.	
DR	Novel purified hepatitis C virus polypeptide comprising epitope encoded by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C virus.	
PT	Example 1; Fig 17; 79pp; English.	
XX	The present invention relates to hepatitis C virus (HCV) proteins and cDNA sequences. The sequences are useful in immunoassays for detecting antibodies directed against HCV antigen; preparing host cells transformed with a recombinant polynucleotide; screening antiviral agents and	

XX	AA818541;	
AC	15-JAN-2001	(first entry)
DT	Polyprotein encoded by sense strand of HCV.	
XX	Hepatitis C virus; HCV; antisense polynucleotide; polyprotein; viral infectivity; viral replication.	
DE	Hepatitis C virus.	
KW	EP1034785-A2.	
KW	13-SEP-2000.	
XX	16-MAR-1990; 2000EP-00109602.	
XX	17-MAR-1989; 89US-00325338.	
PR	20-APR-1989; 89US-00341334.	
PR	18-MAY-1989; 89US-00355002.	
PR	16-MAR-1990; 90EP-00302866.	
XX	(CHIR ) CHIRON CORP.	
PA	Houghton M, Choo Q, Kuo G;	
PI	WPI; 2000-566891/53.	
XX	N-PSDB; AAA75297.	
DR	Novel composition comprising a hepatitis C virus antisense polynucleotide which is complementary to or corresponds to a sense strand of the virus genome, and selectively hybridizes to it.	
PT	Example; Fig 17; 75pp; English.	
XX	The specification describes a pharmaceutical composition which comprises a hepatitis C virus (HCV) antisense polynucleotide. The HCV is characterized by a positive stranded RNA genome which has 40% homology at the polypeptide level to a HCV polypeptide. The antisense polynucleotide binds to cellular polynucleotides which enhance and/or are required for viral infectivity, replicative ability or chronicity. The antisense polynucleotides may also be designed to bind with high specificity, to be of increased stability, to be stable and to have low toxicity. The composition also comprises an agent which causes viral RNA to be inactive. The composition is used for preventing HCV replication in a system. The present sequence is encoded by a novel HCV cDNA sequence, which is used in the course of the invention	
CC	Sequence 2955 AA;	
SQ	Query Match	99.5%; Score 3602; DB 3; Length 2955;
	Best Local Similarity	99.6%; Pred. No. 6.6e-304;
	Matches 683; Conservative	1; Mismatches 2; Indels 0; Gaps 0;
QY	1	MAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTPLATCINGCVWTVYHGA 60
DB	1026	LAPITAAQOTRGLGCIITSLTGRDNQVEGEVQIVSTAAQTPLATCINGCVWTVYHGA 1085
QY	61	GTRTIASPKGPVIQMYTNVDQDLVGMPPAQSGRSRLTCTCGSSDLYLVRHADVIPVRRR 120
DB	1086	GTRTIASPKGPVIQMYTNVDQDLVGMPPAQSGRSRLTCTCGSSDLYLVRHADVIPVRRR 1145
QY	121	GDSRGSLLSPRPIISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180
DB	1146	GDSRGSLLSPRPIISYLGSGGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205
QY	181	RSPVFTDNSSPPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGVKVLNPSVAATLGF 240
DB	1206	RSPVFTDNSSPPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGVKVLNPSVAATLGF 1265
QY	241	AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGFADCGCGGGAYDIIICDECHSTDATS 300

CC determining the effect of antiviral agent in inhibiting viral replication  
CC in cell culture system; and developing vaccine for treating HCV  
CC infection.

XX Sequence 2955 AA;

Query Match 99.5%; Score 3602; DB 8; Length 2955;  
Best Local Similarity 99.6%; Pred. No. 6.6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Db :  
1026 LAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 1085  
Qy 61 GTRTIASPKGPVIOQMTNVQDLVWGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db :  
1086 GTRTIASPKGPVIOQMTNVQDLVWGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db :  
1146 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 240  
Db :  
1206 RSPVFTDSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDATS 300  
Db :  
1266 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDOAETAGARLVVLATAPPGSVTVPHNPTEEVALSTTGSIPFYGKAIPLEVI 360  
Db :  
1326 ILGIGTVLDOAETAGARLVVLATAPPGSVTVPHNPTEEVALSTTGSIPFYGKAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKCDLAKLVAGINAVAYRGLDVSVIPPTGDDVVVATDALMTGYT 420  
Db :  
1386 KGRHLIFCHSKKCDLAKLVAGINAVAYRGLDVSVIPPTGDDVVVATDALMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRRGRTGRKPGIYRFVAPG 480  
Db :  
1446 GDFDSVIDCNTCTVTQTVDFSLDPTFTTITLPODAVSRQRRGRTGRKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTGFLPVCODHLEFEGVPTGL 540  
Db :  
1506 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTGFLPVCODHLEFEGVPTGL 1565  
Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMWKCLIRLKPTLHGPTPLL 600  
Db :  
1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMWKCLIRLKPTLHGPTPLL 1625  
Qy 601 YRLGAVONEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCWTVIGR 660  
Db :  
1626 YRLGAVONEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCWTVIGR 1685  
Qy 661 VVLGKPAIIPDREVLRYREFDEMEEC 686  
Db :  
1686 VVLGKPAIIPDREVLRYREFDEMEEC 1711

RESULT 14

AAR90931  
ID AAR90931 standard; protein; 3011 AA.

XX AAR90931;

AC 25-MAR-2003 (revised)

DT 15-MAY-1996 (first entry)

XX Hepatitis C virus polyprotein.

XX Non-A non-B hepatitis virus; NANBHV; HCV; antigen; detection; diagnosis;  
antibodies.

OS Hepatitis C virus.

XX Key Location/Qualifiers  
FH Misc-difference 1..122 /label= antigen  
FT /note= "C22; AAR90936"  
FT Misc-difference 199..328 /label= antigen  
FT /note= "S2; AAR90935"  
FT Misc-difference 1192..1457 /label= antigen  
FT /note= "C33C; AAR90932"  
FT Misc-difference 1569..1931 /label= antigen  
FT /note= "C100; AAR90933"  
FT Misc-difference 2054..2464 /label= antigen  
FT /note= "NS5; AAR90934"

XX EP693687-A1.

PN 24-JAN-1996.

XX 03-APR-1991; 95EP-00114016.

XX 04-APR-1990; 90US-00504352.

XX (CHIR ) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 1996-117956/13.

XX N-PSDB; AAT12710.

XX Combinations of synthetic Hepatitis C Virus antigens - provide more effective diagnosis of Non-A, Non-B Hepatitis.

XX Disclosure; Fig 1(A-Y); 53pp; English.

XX The combination comprises an HCV antigen from the C domain (pref. C22 - AAR90936) and at least one HCV antigen from the NS3 (pref. C33C - AAR90932), NS4 (pref. C100 - AAR90933), S (pref. S2 - AAR90935) or NS5 (AAR90934) domain. The antigens may in the form of a fusion protein, a simple physical mixture, or the individual antigens commonly bound to a solid matrix. They are pref. prepd. by recombinant DNA techniques (primers are given in AAT12711-T12716), but can be synthesised or isolated from HCV using affinity chromatography. (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 3011 AA;

Query Match 99.5%; Score 3602; DB 2; Length 3011;

Best Local Similarity 99.6%; Pred. No. 6.6e-304;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 60  
Db :  
1026 LAPITAYAAQQTGRLGCGIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA 1085

Qy 61 GTRTIASPKGPVIOQMTNVQDLVWGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db :  
1086 GTRTIASPKGPVIOQMTNVQDLVWGPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db :  
1146 GDSRGSLLSPRPISYLGSGGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 240  
Db :  
1206 RSPVFTDSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAQYKVLVLPNSVAATLGFG 1265

Qy 241 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDATS 300  
Db :  
1266 AYMSKAHGIDPNRTGVRTITTTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDATS 1325

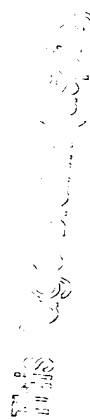
Db	1266	AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGCGSGAVDIIICDECHSTDATS	1325
Qy	301	ILIGITVLDOAETAGARLVVLATATPPGVTVPHPNIEEVALSTTGEIPFYKAIPLEVI	360
Db	1326	ILIGITVLDOAETAGARLVVLATATPPGVTVPHPNIEEVALSTTGEIPFYKAIPLEVI	1385
Qy	361	KGGRHLIFCHSKKKKCDLAAKLVALGINAVAYYRGLDVSVIPIGDVVVVATDALMTGYT	420
Db	1386	KGGRHLIFCHSKKKKCDLAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVVATDALMTGYT	1445
Qy	421	GFDSVIDCNTCTVTQVDFSLDPTFTTITLPPQDAVSRTQRRGTRGKPGIYRFVAPG	480
Db	1446	GFDSVIDCNTCTVTQVDFSLDPTFTTITLPPQDAVSRTQRRGTRGKPGIYRFVAPG	1505
Qy	481	ERPSGMFSSVLCECYDAGCAWYELTPAETVRLRAYMNTPLPVQCDHLEFWEGVFTGL	540
Db	1506	ERPSGMFSSVLCECYDAGCAWYELTPAETVRLRAYMNTPLPVQCDHLEFWEGVFTGL	1565
Qy	541	THIDHFLSQTQSGENLPVLVAYQATVCARAQPPSWDQWKCLIRLKLHGTPLLL	600
Db	1566	THIDHFLSQTQSGENLPVLVAYQATVCARAQPPSWDQWKCLIRLKLHGTPLLL	1625
Qy	601	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGR	660
Db	1626	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGR	1685
Qy	661	VVLGKPAIIPREVLYRFEDEMEEC	686
Db	1686	VVLGKPAIIPREVLYRFEDEMEEC	1711
RESULT 15			
AAW34480	ID	AAW34480 standard; protein; 3011 AA.	
XX	AC	AAW34480;	
XX	AC	AAW34480;	
XX	AC	AAW34480;	
DT	25-MAR-2003	(revised)	
DT	16-MAR-1998	(first entry)	
DE	HCV polyprotein.		
XX	PCR primer; amplify; HCV; hepatitis c virus; antigen combination; NS3;		
KW	C domain; S domain; NS5; HCV polyprotein; anti-HCV antibody; detection;		
KW	NS4.		
XX	Hepatitis C virus.		
OS	Key	Location/Qualifiers	
FT	Misc-difference	366 /note= "can optionally be Arg"	
FT	Misc-difference	372 /note= "can optionally be Thr"	
FT	Misc-difference	867 /note= "can optionally be Thr"	
FT	Misc-difference	1341 /note= "can optionally be Val"	
FT	Misc-difference	2148 /note= "can optionally be Ile"	
FT	Misc-difference	2883 /note= "can optionally be Asn"	
FT	Misc-difference	3681 /note= "can optionally be Ser"	
FT	Misc-difference	3690 /note= "can optionally be Thr"	
FT	Misc-difference	4167 /note= "can optionally be Leu"	
FT	Misc-difference	4323 /note= "can optionally be Val"	
FT	Misc-difference	4701 /note= "can optionally be Tyr"	
FT	Misc-difference	4752 /note= "can optionally be Ser"	

FT	Misc-difference	5970 /note= "can optionally be Gly"	
FT	Misc-difference	6183 /note= "can optionally be His"	
FT	Misc-difference	6186 /note= "can optionally be Cys"	
FT	Misc-difference	6402 /note= "can optionally be Val"	
FT	Misc-difference	7386 /note= "can optionally be Ser"	
FT	Misc-difference	7494 /note= "can optionally be Phe"	
FT	Misc-difference	7497 /note= "can optionally be Ala"	
FT	Misc-difference	7845 /note= "can optionally be Phe"	
FT	Misc-difference	8409 /note= "can optionally be Gly"	
FT	Misc-difference	9102 /note= "can optionally be Gly"	
FT	Misc-difference	9327 /note= "can optionally be Pro"	
XX	US5683864-A.		
PN	04-NOV-1997.		
XX	07-JUL-1992;	92US-00910760.	
XX	18-NOV-1987;	87US-00122714.	
PR	30-DEC-1987;	87US-00139886.	
PR	26-FEB-1988;	88US-00161072.	
PR	06-MAY-1988;	88US-00191263.	
PR	26-OCT-1988;	88US-00263584.	
PR	14-NOV-1988;	88US-00271450.	
PR	17-MAR-1989;	89US-00325338.	
PR	20-APR-1989;	89US-00341334.	
PR	21-APR-1989;	89US-00353896.	
PR	18-MAY-1989;	89US-00355002.	
PR	04-APR-1990;	90US-00504352.	
XX	(CHIR ) CHIRON CORP.		
XX	Kuo G, Houghton M, Choo Q;		
PI	WPI; 1997-548976/50.		
DR	N-PSDB; AAT99981.		
DR	Combination of three hepatitis C virus antigens - used for detection of specific antibodies to diagnose infection.		
XX	Disclosure; Col 25-46; 57pp; English.		
PS	This sequence represents the Hepatitis C virus polyprotein. Fragments of the DNA encoding this sequence can be amplified and used in the combination of HCV antigens of the invention. The HCV antigen combination comprises an antigen (Ag1) comprising the C domain (i.e. amino acids (aa) 1-120 of the HCV polyprotein), or its immunologically reactive fragment containing at least 8 aa. It also comprises two additional antigens from two different polyprotein domains, including at least 8 aa from the NS3, NS4, S or NS5 domains of the polyprotein, corresponding, respectively, to aa 1050-1640; 1640-2000; 120-400 and 2000-3011 of the HCV polyprotein. Alternatively, Ag1 contains at least 8 aa from the 1-122 or 9-177 aa regions of the HCV polyprotein. These antigen combinations are used diagnostically to detect anti-HCV antibodies, using any standard immunoassay format. These antigen combinations have a broader range of reactivity with antibodies than any antigen individually. (Updated on 25-MAR-2003 to correct PR field.)		
XX	Sequence 3011 AA;		
SQ	Query Match	99.5%;	Score 3602; DB 2; Length 3011;
	Best Local Similarity	99.6%;	Pred. No. 6.8e-304;



Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60
Db	:
Qy	1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085
Db	:
Qy	61 GTRTIASPKGPVIOMYTNVDODLVGWPAPOGSRSLTCTCGSSDLYLVTTRHADVIPVRRR 120
Db	:
Qy	1086 GTRTIASPKGPVIOMYTNVDODLVGWPAPOGSRSLTCTCGSSDLYLVTTRHADVIPVRRR 1145
Db	:
Qy	121 GDSRGSLLSPRPISYLAGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db	:
Qy	1146 GDSRGSLLSPRPISYLAGSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205
Db	:
Qy	181 RSPVFTDNSPPVPVQSFQVAHLHAPTCSGKSTKVPAAAYAAQGYKVLVLPNPSVAATLGFG 240
Db	:
Qy	1206 RSPVFTDNSPPVPVQSFQVAHLHAPTCSGKSTKVPAAAYAAQGYKVLVLPNPSVAATLGFG 1265
Db	:
Qy	241 AYMSKAHGIDPNIRHTGVRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSTDATS 300
Db	:
Qy	1266 AYMSKAHGIDPNIRHTGVRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSTDATS 1325
Db	:
Qy	301 ILGIGTVLDOAETAGARLVLATATPPGSVTVPHPNIEEVALSTTGEIPFVGKAI PLEVI 360
Db	:
Qy	1326 ILGIGTVLDOAETAGARLVLATATPPGSVTVPHPNIEEVALSTTGEIPFVGKAI PLEVI 1385
Db	:
Qy	361 KGRHLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVATDALMTGYT 420
Db	:
Qy	1386 KGRHLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVATDALMTGYT 1445
Db	:
Qy	421 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRRTORRGRTGRKPGIYRFVAPG 480
Db	:
Qy	1446 GDFDSVIDCNTCVTQTVDFSLDPTFTTETITLPQDAVSRRTORRGRTGRKPGIYRFVAPG 1505
Db	:
Qy	481 ERPSGMFDSVLCFCYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFNEGVTGL 540
Db	:
Qy	1506 ERPSGMFDSVLCFCYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFNEGVTGL 1565
Db	:
Qy	541 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPTRLHGPTPLL 600
Db	:
Qy	1566 THIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQMKCLIRLKPTRLHGPTPLL 1625
Db	:
Qy	601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVI VGR 660
Db	:
Qy	1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCWVI VGR 1685
Db	:
Qy	661 VVLSGKPAIIPDREVLYREFDEMEEC 686
Db	:
Qy	1686 VVLSGKPAIIPDREVLYREFDEMEEC 1711

Search completed: November 7, 2005, 20:09:55  
Job time : 84.2548 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:00:21 ; Search time 14.1714 Seconds  
(without alignments)  
4657.604 Million cell updates/sec

Title: US-10-658-782-2  
Perfect score: 3619  
Sequence: 1 MAPITAYAAQTGRLGCIIT.....PAIIPREVLYREFDEMEEC 686

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: Pir1:.\*  
2: Pir2:.\*  
3: Pir3:.\*  
4: Pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3602	99.5	3011	1 GNVWC3	genome polyprotein
2	3542	97.9	3011	1 S40770	genome polyprotein
3	3513	97.1	3011	1 GNVVTC	genome polyprotein
4	3426	94.7	3010	1 GNVVTC	genome polyprotein
5	3417	94.4	3010	1 A45573	genome polyprotein
6	3408	94.2	3010	1 GNVWCJ	genome polyprotein
7	3402	94.0	3010	1 GNVVTW	genome polyprotein
8	3375	93.3	3010	1 S18030	genome polyprotein
9	3082	85.2	3014	1 JC5620	genome polyprotein
10	2989	82.6	3033	1 JQ1303	genome polyprotein
11	2979	82.3	3033	1 GNVVJ8	genome polyprotein
12	1993	55.1	876	2 PC2219	polyprotein - hepa
13	1972.5	54.5	386	2 S68016	ATPase/RNA helicase
14	1388	38.4	492	2 PS0326	polyprotein - hepa
15	1383	38.2	716	2 JQ1366	polyprotein - hepa
16	1101	30.4	3005	2 T08841	polyprotein - dour
17	1075	29.7	2970	2 T08839	polyprotein - maru
18	1055	29.2	216	2 S21337	genome polyprotein
19	1001	27.7	194	2 S06067	nonstructural prot
20	998	27.6	194	2 A54317	probable nonstruct
21	980	27.1	182	2 S32748	genome polyprotein
22	947	26.2	184	2 A61196	genome polyprotein
23	728	20.1	209	2 PC1306	genome polyprotein
24	720	19.9	135	2 PS0327	polyprotein - hepa
25	719	19.9	135	2 PS0328	polyprotein - hepa
26	693	19.1	135	2 PC1309	polyprotein - hepa
27	397	11.0	142	2 PC1307	genome polyprotein
28	310	8.6	102	2 PC6028	genome polyprotein
29	280	7.7	132	2 PQ0394	genome polyprotein

30	280	7.7	132	2	PQ0396	genome polyprotein
31	277	7.7	157	2	PQ0401	genome polyprotein
32	276	7.6	125	2	S35629	hypothetical prote
33	273	7.5	3898	1	GNVWHC	genome polyprotein
34	273	7.5	3898	1	GNVWHB	genome polyprotein
35	271	7.5	3898	2	S57437	genome polyprotein
36	268.5	7.4	3988	1	GNVWEV	genome polyprotein
37	267.5	7.4	3898	1	A44217	genome polyprotein
38	265.5	7.3	266	2	PQ0393	polyprotein - hog
39	264	7.3	3898	2	S58295	genome polyprotein
40	227.5	6.3	3341	1	A42996	genome polyprotein
41	223	6.2	3434	1	GNVWV	genome polyprotein
42	202	5.6	3125	1	GNVSPP	genome polyprotein
43	202	5.6	3140	2	S47508	genome polyprotein
44	201	5.6	3140	1	GNVSRA	genome polyprotein
45	201	5.6	3141	1	GNVSPD	genome polyprotein

ALIGNMENTS

RESULT 1

GNVWC3

genome polyprotein - hepatitis C virus (strain HCV-1)  
N:Contains: capsid protein C; envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A39166; PQ0403; PQ0404  
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A:Title: Genetic organization and diversity of the hepatitis C virus.  
A:Reference number: A39166; MUID:91172826; PMID:1848704  
A:Accession: A39166  
A:Molecule type: mRNA  
A:Residues: 1-3011 <CHO>  
A:CROSS-references: UNIPROT:P26664; GB:M62321; NID:9329873; PIDN:AAA45676.1; PID:9329874  
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992  
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e A:Reference number: PQ0393; MUID:92268871; PMID:1316939  
A:Accession: PQ0403  
A:Molecule type: Genomic RNA  
A:Residues: 1577-1633 <CHA>  
A:CROSS-references: DDBJ:D10128  
A:Experimental source: isolates E-b16  
A:Accession: PQ0404  
A>Status: Preliminary  
A:Molecule type: Genomic RNA  
A:Residues: 1577-1633 <CH2>  
A:Experimental source: isolates E-b17  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F:115-191/Product: major envelope protein M #status predicted <CPC>  
F:116-191/Product: capsid protein C #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NAA>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 99.5%; Score 3602; DB 1; Length 3011;  
Best Local Similarity 99.6%; Pred. No. 7.4e-240;  
Matches 693; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPITAYAAQTGRLGCIITSTGRDKNOVEGVQIVSTAAQTFLATCINGVCTVYHCA 60  
:|||||

Db 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLVLTTRHADVIPVRRR 120  
Db 1086 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLVLTTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDSSPPVQSFQVAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDSSPPVQSFQVAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDAQETAGARLVLTATPPGSGVTVPHPNIEEVALSTTGIBPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDAQETAGARLVLTATPPGSGVTVPHPNIEEVALSTTGIBPFYGKAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 600  
Db 1566 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 1625  
Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAYCLSTGCWVIVGR 660  
Db 1626 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAYCLSTGCWVIVGR 1685  
Qy 661 VLSGKPAIIPDREVLRYREFDEMEEC 686  
Db 1686 VLSGKPAIIPDREVLRYREFDEMEEC 1711  
RESULT 2  
S40770  
genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: S40770; PC1285  
R;Okamoto, H.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S40770  
A;Accession: S40770  
A;Molecule type: genomic RNA  
A;Residues: 1-3011 <OKA>  
R;Cross-references: UNIPROT:Q03463; EMBL:D10749; NID:G221586; PIDN:BAA01582.1; PID:G2215  
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, Jpn. J. Exp. Med. 60, 167-177, 1990  
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A;Reference number: PC1284; MUID:91013116; PMID:2170712  
A;Accession: PC1285  
A;Molecule type: genomic RNA  
A;Residues: 1-513 <OK2>  
A;Cross-references: GB:D00831; NID:G221511; PIDN:BAA00705.1; PID:G221512  
A;Experimental source: isolate HC-J1  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <BPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
Query Match 97.9% Score 3542; DB 1; Length 3011;  
Best Local Similarity 96.9%; Pred. No. 1e-235;  
Matches 665; Conservative 10; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 MAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db 1026 LAPITAYAAQTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLVLTTRHADVIPVRRR 120  
Db 1086 GTRTIASPKGPVIQMYTNVDQDLVGPAPQGSRSITPCTCGSSDLVLTTRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSSGGLLCPAGHAGVGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDSSPPVQSFQVAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDSSPPVQSFQVAHLHAPTSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYKGLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDAQETAGARLVLTATPPGSGVTVPHPNIEEVALSTTGIBPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDAQETAGARLVLTATPPGSGVTVPHPNIEEVALSTTGIBPFYGKAIPLEVI 1385  
Qy 361 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKDELAALKVALGINAVAYRGLDVSVIPPIGDVVVATDALTMTGYT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 540  
Db 1506 ERPSGMFDSVLCBECYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 600  
Db 1566 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMWKCLIRLKPTLHGPTPL 1625  
Qy 601 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAYCLSTGCWVIVGR 660  
Db 1626 YRLGAVONEITLTHPVTKYIMTCSADLEVVTSWLVGGVLAALAYCLSTGCWVIVGR 1685  
Qy 661 VLSGKPAIIPDREVLRYREFDEMEEC 686  
Db 1686 VLSGKPAIIPDREVLRYREFDEMEEC 1711  
RESULT 3  
GNVWCH  
genome polyprotein - hepatitis C virus (strain H)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C;Accession: A36814; A41546  
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
submitted to GenBank, July 1992  
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992  
A;Reference number: A36814  
A;Accession: A36814  
A;Molecule type: genomic RNA  
A;Residues: 1-3011 <INP>  
A;Cross-references: UNIPROT:P27958; GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
R;Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: compared to GenBank, July 1992  
A;Reference number: A41546; MUID:92052256; PMID:1658800  
A;Contents: annotation  
A;Note: neither amino acid nor nucleotide sequence is given  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1319/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match 97.1%; Score 3513; DB 1; Length 3011;  
Best Local Similarity 96.6%; Pred. No. 1e-233;  
Matches 663; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 60  
Db 1026 LAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 1085

Qy 61 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGLSPRISYLYKSGSGGLLCPGTHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 1265

Qy 241 AYMSKAHGIDNIRGVTITTTGSPITVSTYTKFLADGCGSGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDNIRGVTITTTGSPITVSTYTKFLADGCGSGAYDIIICDECHSTDATS 1325

Qy 301 ILIGTGLVLDQAETAGARLVWLATATPPGVTVPHPNIEVALSTTGEIPFYKAIPLV 360  
Db 1326 ILIGTGLVLDQAETAGARLVWLATATPPGVTVPHPNIEVALSTTGEIPFYKAIPLV 1385

Qy 361 KGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVVVATDALTMTGYT 420  
Db 1386 KGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVVVATDALTMTGYT 1445

Qy 421 GDFSDVIDCNTVQTVDPSLDPTTETITLPODAVSRTOGRGKPGIYRFVAPG 480  
Db 1446 GDFSDVIDCNTVQTVDPSLDPTTETITLPODAVSRTOGRGKPGIYRFVAPG 1505

Qy 481 BRPSGMFDSVLCCEYDAGCAWELTPAETTVRLRAYMNTPLPVQCQDLHEFWGVTGL 540  
Db 1506 BRPSGMFDSVLCCEYDAGCAWELTPAETTVRLRAYMNTPLPVQCQDLHEFWGVTGL 1565

Qy 541 THIDAHFLSQTQSGENLPYLVAQATVCAQAQAPPPSWDMWKCLIRLKTPLHGPTPLL 600

Db 1566 THIDAHFLSQTQSGENLPYLVAQATVCAQAQAPPPSWDMWKCLIRLKTPLHGPTPLL 1625

Qy 601 YRLGAVQNEITLTHPTVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCVVIGR 660

Db 1626 YRLGAVQNEITLTHPTVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCVVIGR 1685

Qy 661 VVLGKPAIIPDRVLYREFDEMEEC 686

Db 1686 IVLSGKPAIIPDRVLYREFDEMEEC 1711

RESULT 4  
GNMVT  
genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: A38465  
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; J. Virol. 65, 1105-1113, 1991  
A;Title: Structure and organization of the hepatitis C virus genome isolated from human  
A;Reference number: A38465; MUID:91140698; PMID:1847440  
A;Accession: A38465  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <TAK>  
A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G32977  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus genome polyprotein  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1319/Region: DEXH motif  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,234,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 94.7%; Score 3426; DB 1; Length 3010;  
Best Local Similarity 92.0%; Pred. No. 1e-227;  
Matches 631; Conservative 37; Mismatches 18; Indels 0; Gaps 0;

Qy 1 MAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 60  
Db 1026 LAPITAYAAQOTRGLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCTWTVYHGA 1085

Qy 61 GTRTTASPKGPVIOYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 120  
Db 1086 GSKTLAAPGPIQYNTVDQDLVGPAPQGSRSITPTCTCGSSDLYLVTRHADVIPVRRR 1145

Qy 121 GDSRGLSPRISYLYKSGSGGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGLSPRISYLYKSGSGGLLCPGTHAGVIFRAAVCTRGVAKAVDFIPVESMETM 1205

Qy 181 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPPQSFQVAHLHAPTGSKSTKVPAAAYAGYKVLVLPNSVAATLGF 1265

Qy 241 AYMSKAHGIDNIRGVTITTTGSPITVSTYTKFLADGCGSGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDNIRGVTITTTGSPITVSTYTKFLADGCGSGAYDIIICDECHSTDATS 1325

Qy 301 ILIGTGLVLDQAETAGARLVWLATATPPGVTVPHPNIEVALSTTGEIPFYKAIPLV 360  
Db 1326 ILIGTGLVLDQAETAGARLVWLATATPPGVTVPHPNIEVALSTTGEIPFYKAIPIEAI 1385

Qy 361 KGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVVVATDALTMTGYT 420

Db 1386 RGRHLLFCHSKKCDLAALKSLGLGNVAYAYRGLDVSVPITIGDVVVATDALMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVQTQTVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVAPG 480  
Db 1446 GDFDSVIDCNTCTVQTQTVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVTPG 1505  
Qy 481 ERSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQAPPPSWDMWKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPPSWDMWKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCISLTCGCVWIVGR 660  
Db 1626 YRLGAVQNEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCISLTCGCVWIVGR 1685  
Qy 661 VVLGSKPAIIPDREVLRYREDEMEEC 686  
Db 1686 IILSGRPAIVPDRELLYQSFDEMEEC 1711  
RESULT 5  
A45573  
genome polyprotein - hepatitis C virus (strain JT)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome from a single Japanese carrier: s  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: A45573  
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, T.; Shimotohno, K.  
Virus Res. 23, 39-53, 1992  
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s  
A;Reference number: A45573; MUID:92295714; PMID:1318627  
A;Accession: A45573  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-3010 <TAN>  
A;Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1;  
A;Experimental source: HCV-JT  
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome from Japanese patients v  
F;1220-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
Query Match 94.4%; Score 3417; DB 1; Length 3010;  
Best Local Similarity 92.0%; Pred. No. 4.3e-217;  
Matches 631; Conservative 34; Mismatches 22; Indels 0; Gaps 0;  
Qy 1 MAPTAAQOTRGLGCIITSLGRDNQVEGEVQIVSTAQTPLATCINGCVCHTVVHGA 60  
Db 1026 LAPITAAQOTRGLGCIITSLGRDNQVEGEVQIVSTAQTPLATCINGCVCHTVVHGA 1085  
Qy 61 GTRTASPKPGVIOQMTYNTVDQLVGPWAPQSGRSRLTCTCGSSDLYLTVRHADVIPVRRR 120  
Db 1086 GSXTLAGPKGPITQMTYNTVDQLVGPWAPQSGRSRLTCTCGSSDLYLTVRHADVIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDGRGSLLSPRPISYLVKSGSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205

Qy 181 RSPVFTDNSSPPVPOSFOVAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 240  
Db 1206 RSPVFTDNSSPPVPOSFOVAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGIDPNIRTTGSPITYSTYKFLADGGCGSGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTTGSPITYSTYKFLADGGCGSGGAYDIIICDECHSTDTST 1325  
Qy 301 ILGITVLDQAEATAGARLVVLTATATPPGSVTVPHNIEEVALSTTGPYKKAIPLEVI 360  
Db 1326 ILGITVLDQAEATAGARLVVLTATATPPGSVTVPHNIEEVALSTTGPYKKAIPLEAI 1385  
Qy 361 KGRHLLFCHSKKCDLAALKSLGLGNVAYAYRGLDVSVPITIGDVVVATDALMTGYT 420  
Db 1386 KGRHLLFCHSKKCDLAALKSLGLGNVAYAYRGLDVSVPITIGDVVVATDALMTGYT 1445  
Qy 421 GDFDSVIDCNTCTVQTQTVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVAPG 480  
Db 1446 GDFDSVIDCNTCTVQTQTVDFSLDPTFTTETITLPODAVSRQRRGRTGRGPGIYRFPVTPG 1505  
Qy 481 ERSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 540  
Db 1506 ERSGMFDSSVLCBCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEVFTGL 1565  
Qy 541 THIDAHFLSOTKSGENLPYLVAQATVCARAQAPPPSWDMWKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSOTKQAGDNFPYLVAQATVCARAQAPPPSWDMWKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCISLTCGCVWIVGR 660  
Db 1626 YRLGAVQNEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCISLTCGCVWIVGR 1685  
Qy 661 VVLGSKPAIIPDREVLRYREDEMEEC 686  
Db 1686 IILSGRPAIVPDRELLYQSFDEMEEC 1711  
RESULT 6  
GNWVCJ  
genome polyprotein - hepatitis C virus (strain J)  
N;Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: A39253; PS0086  
R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990  
A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients v  
A;Reference number: A39253; MUID:91088550; PMID:2175903  
A;Accession: A39253  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <KAT>  
A;Cross-references: UNIPROT:P26662; GB:D90208; NID:g221610; PIDN:BAA14233.1; PID:g221611  
R;Kato, N.; Ohkoshi, S.; Shimotohno, K.  
Proc. Jpn. Acad. 65B, 219-223, 1989  
A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence varia  
A;Reference number: PS0086  
A;Accession: PS0086  
A;Molecule type: genomic RNA  
A;Residues: 2650-2707 <KAT>  
A;Experimental source: Japanese isolate  
A;Comment: The cleavage sites of this polyprotein have not been determined.  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome from Japanese patients v  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2  
Query Match 94.2%; Score 3408; DB 1; Length 3010;  
Best Local Similarity 91.4%; Pred. No. 1.8e-226;  
Matches 627; Conservative 38; Mismatches 21; Indels 0; Gaps 0;  
Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db 1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIOQMYTNVDQDLVGPAPQGSRLTCTCGSSDLYLVTRHADYIPVRRR 120  
Db 1086 GSKTLAGPKGPITQMYTNVDQDLVGPAPPGARSMTPTCTCGSSDLYLVTRHADYIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSGGPGLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSGGPGLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVPOQFVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDNSSPPVPOQFVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGFG 1255  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDOAETAGARLVLATATPPGCVTVPHNIEEVALSTGIBPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDOAETAGARLVLATATPPGCVTVPHNIEEVALSTGIBPFYGKAIPLEVI 1385  
Qy 361 KGBRHILFCHSKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVATDALTGT 420  
Db 1386 KGBRHILFCHSKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVATDALTGT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTQRRGRTGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTQRRGRTGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYNTPLGLPVCODHLEFEGVFTGL 540  
Db 1506 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYNTPLGLPVCODHLEFEGVFTGL 1565  
Qy 541 THIDAHFLSQTQKAGDNLPLYLVAQVTCARQAAPPSPSDQMKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSQTQKAGDNLPLYLVAQVTCARQAAPPSPSDQMKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 1685  
Qy 661 VVLGSKPAIIPDREVLYREFDEMESEC 686  
Db 1686 IILSGKPAVVDREVLYQEFDEMESEC 1711

RESULT 7  
GNVVTW  
Genome polyprotein - hepatitis C virus (strain Taiwan)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain Taiwan)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C:Accession: A40244  
R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.  
Virology 189, 102-113, 1992  
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the  
A:Reference number: A40244; MUID:92230206; PMID:1314449  
A:Accession: A40244

A:Molecule type: genomic RNA  
A:Residues: 1-3010 <CHE>  
A:Cross-references: UNIPROT:P29846; GB:M84754  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F;1-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196, 209, 233, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 207  
Query Match 94.0%; Score 3402; DB 1; Length 3010;  
Best Local Similarity 91.4%; Pred. No. 4.7e-226;  
Matches 627; Conservative 38; Mismatches 21; Indels 0; Gaps 0;  
Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db 1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGPVIOQMYTNVDQDLVGPAPQGSRLTCTCGSSDLYLVTRHADYIPVRRR 120  
Db 1086 GSKTLAGPKGPITQMYTNVDQDLVGPAPQGSRLTCTCGSSDLYLVTRHADYIPVRRR 1145  
Qy 121 GDSRGSLLSPRPISYLGKSGGPGLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db 1146 GDSRGSLLSPRPISYLGKSGGPGLLCPAGHAGTGFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVPOQFVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGFG 240  
Db 1206 RSPVFTDNSSPPVPOQFVAHLHAPTGSKSTKVPAAQAQGYKVLVLPNSVAATLGFG 1255  
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATS 300  
Db 1266 AYMSKAHGIDPNIRTVGRTITTTGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATS 1325  
Qy 301 ILGIGTVLDOAETAGARLVLATATPPGCVTVPHNIEEVALSTGIBPFYGKAIPLEVI 360  
Db 1326 ILGIGTVLDOAETAGARLVLATATPPGCVTVPHNIEEVALSTGIBPFYGKAIPLEVI 1385  
Qy 361 KGBRHILFCHSKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVATDALTGT 420  
Db 1386 KGBRHILFCHSKKKDELAALVALGINAVAYRGLDVSVIPPIGDVVVATDALTGT 1445  
Qy 421 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTQRRGRTGKPGIYRFVAPG 480  
Db 1446 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTQRRGRTGKPGIYRFVAPG 1505  
Qy 481 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYNTPLGLPVCODHLEFEGVFTGL 540  
Db 1506 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYNTPLGLPVCODHLEFEGVFTGL 1565  
Qy 541 THIDAHFLSQTQKAGDNLPLYLVAQVTCARQAAPPSPSDQMKCLIRLKLPTLHGPTPL 600  
Db 1566 THIDAHFLSQTQKAGDNLPLYLVAQVTCARQAAPPSPSDQMKCLIRLKLPTLHGPTPL 1625  
Qy 601 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 660  
Db 1626 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSTWVLVGGVLAALAAAYCLSTGCWVIYGR 1685  
Qy 661 VVLGSKPAIIPDREVLYREFDEMESEC 686  
Db 1686 IILSGKPAVVDREVLYQEFDEMESEC 1711

S18030 genome polyprotein - hepatitis C virus (isolate JX1)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate JX1) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Variety: isolate JX1  
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: S18030; S33570; A48332; S18029  
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
A:Reference number: S18028  
A:Accession: S18030  
A:Molecule type: Genomic RNA  
A:Residues: 1-3010 <HON>  
A:Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
A:Experimental source: Isolate JX1 from an individual  
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
A:Reference number: A48332; MUID:93119270; PMID:8360322  
A:Accession: S33570  
A:Molecule type: Genomic RNA  
A:Residues: 1-547, 'T', '549-621, 'V', '623-624, 'S', '626-652, 'DL', '655-761, 'T', '763-782 <HON>  
A:Cross-references: EMBL:X61591  
A:Note: this sequence is inconsistent with the nucleotide translation  
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Amp; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus polyprotein NS3 #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196, 209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate (A)

Query\_Match 93.3%; Score 3375; DB 1; Length 3010;  
Best Local Similarity 91.3%; Pred. No. 3.4e-244;  
Matches 626; Conservative 33; Mismatches 27; Indels 0; Gaps 0;

Qy	1	M A P I T A A Q Q R G L G C I I T S L T G R D K N O V E G E V Q I V S T A A Q T F L A T C I N G V C W T V Y H G A	60
Db	1026	L A P I T A S Q Q T R G L F G C I V T S L T G R D K N O V E G E A Q V S T A T Q S F L A T C V G C W T V Y H G A	1085
Qy	61	G T R T I A S P K G P V I Q M Y T N V D Q D L V G W P A P Q S R S L T P C T G S S D L Y L V T R H A D V I P V R R R	120
Db	1086	G S K T L A G K P G I N Q M Y T N V D Q D L V G W Q A P S A A S L T P C T Y G S S D L Y L V T R H A D V I P V R R R	1145
Qy	121	G D S R G S L L S P R P I S Y L K G S S G G P L L C P A G H A V G I F R A A V C T R G Y A K A V D F I P V E N L E T T M	180
Db	1146	G D S R G S L L S P R P S Y L K G S S G G P L L C P S G H A V G I F R A A V C T R G Y A K A V D F I I P V E S M E T T M	1205
Qy	181	R S P V F T D N S S P V P V Q S F Q V A H L H A P T G S G K S T K V P A A Y A A Q G K V L V L N P S V A A T L G F G	240
Db	1206	R S P V F T D N S S P A N P Q T F Q V A H L H A P T G S G K S T K V P A A Y A A Q G K V L V L N P S V A A T L G F G	1265
Qy	241	A Y M S K A H G I D P N I R T G V R T I T T G S P I T Y S T Y K F L A D G C G S G G A Y D I I I C D E C H S T D A T S	300
Db	1266	A Y M S K A H G V D P N I S T G V R T I T T G A P I T Y S T Y K F L A D G C G S G G A Y D I I I C D E C H S T D T S	1325
Qy	301	I L G I G T V L D Q A E T A G A R L V L A T A T P G S V T V P H N I E V A L S T T G E I P F Y G K A I P L E V I	360
Db	1326	I L G I G T V L D Q A E T A G A R L V L A A A T P P G S V T V P H N I E V A L P N T G E I P F Y G K A I P L E T I	1395



```
Db 1147 GDTASLLSPRIISYLGSSGGPIMCPSPGHVGVFRAAVCTRGVAKALEFPVVENLETTM 1206
Qy 181 RSPVFTDNSSPPVPOQSFQVAHLHAPTSKSKTKVPAAYAAQGVKVLVLPSPVAATLGF 240
Db 1207 RSPVFTDNSTPPAPVHEFQVGHLLHAPTSKSKTKVPAAYAAQGVKVLVLPSPVAATLGF 1266
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYSGKFLADGGCGGGAYDIIICDECHSDTATS 300
Db 1267 AYMSRAYGVDPNIRTVGRTITTTGAGITVSTYSGKFFADGGCGGGAYDIIICDECHSDQATT 1326
Qy 301 ILGIGTVLDOAETAGARLVLTATPPGSVTVPHNPNEEVALSTTGEIPFYGKAIPLEVI 360
Db 1327 ILGIGTVLDOAETAGARLVLTATAPPGSVTVPHNPNEEVALPSEGEIPFYGKAIPLEVI 1386
Qy 361 KGGRHLLFCHSKKCDLAALVALGINAVAYRGLDVSVPPIPGDVVVVATDALMTGYT 420
Db 1387 KGGRHLLFANOKKAAKAKETAKONKPGKAVAYRGLDVAVIPATGDDVVVSTDALMTGFT 1446
Qy 421 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTORRGTGRGKPGIYRFPVAPG 480
Db 1447 GDFDSVIDCNSAVTQTVDFSLDPTFTTITTVPQDAVSRSORRGTGRGHIYRYVSSG 1506
Qy 481 ERPSGMFDSVLCBCYDAGCANYELTTPAETTVRLRAYMNTPGLPVCDHLEFMEGVTGL 540
Db 1507 ERPSGIFDSVVLBCYDAGCANYDLTPAETTVRLRAYLNTPLGVCOEHLFEFMEGVTGL 1566
Qy 541 THIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDQMKCLIRLKLPTLHGPTPL 600
Db 1567 TNIDAHMLSQKQGENFPYLVAQVATVCVRAKAPPPSDMTWKMCLIRLKLPTLHGPTPL 1626
Qy 601 YRLGAVQNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
Db 1627 YRLGAVQNEITLTHPTIKYIMACNSADLEVTSTWLVGGVVAALAAAYCLTVGSVAIVGR 1686
Qy 661 VVLSGKAIPITDREVLYREFDEMEEC 686
Db 1687 IILSGREPAITDREVLYQQFDEMEEC 1712

RESULT 10
JQ1303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J6)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: JQ1303
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.
J. Gen. Virol. 72, 2697-2704, 1991
A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A;Reference number: JQ1303; MUID:92044440; PMID:1658196
A;Accession: JQ1303
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26660; GB:D00944; NID:g221650; PIDN:BAA00792.1; PID:g221651
A;Experimental source: isolate HC-J6 from a Japanese individual
A;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane
F;2-115/Product: capsid protein C #status predicted <CP>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F;1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1867-1866/Product: nonstructural protein NS4a #status predicted <NS4a>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 82.6%; Score 2989; DB 1; Length 3033;
```

```
Best Local Similarity 79.9%; Pred. No. 1.5e-197;
Matches 548; Conservative 63; Mismatches 75; Indels 0; Gaps 0;

Qy 1 MAPITAYAQOTRGLGCIITSLTGRDKNQVGEVQIVSTAAQTFLATCINGVCWTVVHGA 60
Db 1030 LAPITAYAQOTRGLGCIITVVMYMTGRDTEQAGEIQVLSTVTSQFLGTTISGLVTVVHGA 1089
Qy 61 GTRTIASPKGPVIMQYTNVDLWGPAPQSGRSRLTPTCTCGSSDLXLYLVTRHADVIPVRR 120
Db 1090 GNKTLAGSRGPVTQMYSSAGDLVGMPSPPGKTSLEPCTCGADVLYLVTRNADVIPARR 1149
Qy 121 GDSRGSLLSRPPTSYLKGSGGGLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLETTM 180
Db 1150 GDXRGALLSRPLSLTLKGSGGGLVCPRGHAVGVFRAAVCSRGVAKSIDIFIPVETLOIVT 1209
Qy 181 RSPVFTDNSSPPVPOQSFQVAHLHAPTSKSKTKVPAAYAAQGVKVLVLPSPVAATLGF 240
Db 1210 RSPVFTDNSTPPAPVHEFQVGHLLHAPTSKSKTKVPAAYAAQGVKVLVLPSPVAATLGF 1269
Qy 241 AYMSKAHGIDPNIRTVGRTITTTGSPITYSTYSGKFLADGGCGGGAYDIIICDECHSDTATS 300
Db 1270 AYLSKAHGIDPNIRTVGRTITTTGAPITYSTYSGKFLADGGCAGGAYDIIICDECHAVDSTT 1329
Qy 301 ILGIGTVLDOAETAGARLVLTATPPGSVTVPHNPNEEVALSTTGEIPFYGKAIPLEVI 360
Db 1330 ILGIGTVLDOAETAGARLVLTATPPGSVTVPHNPNEEVALGQGEIPFYGKAIPLEVI 1389
Qy 361 KGGRHLLFCHSKKCDLAALVALGINAVAYRGLDVSVPPIPGDVVVVATDALMTGYT 420
Db 1390 KGGRHLLFCHSKKCDLAALRGMLNNAVAYRGLDVSVPITQGDVVVATDALMTGFT 1449
Qy 421 GDFDSVIDCNTCVTQTVDVDFSLDPTFTTITLTPQDAVSRTORRGTGRGKPGIYRFPVAPG 480
Db 1450 GDFDSVIDCNAVTVQVDFSLDPTFTTITTVPQDAVSRSORRGTGRGLGIYRYVSTG 1509
Qy 481 ERPSGMFDSVLCBCYDAGCANYELTTPAETTVRLRAYMNTPGLPVCDHLEFMEGVTGL 540
Db 1510 ERASGMFDSVLCBCYDAGCANYELTTPAETTVRLRAYMNTPGLPVCDHLEFMEGVTGL 1569
Qy 541 THIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDQMKCLIRLKLPTLHGPTPL 600
Db 1570 THIDAHFLSQTQKSGENFAYLTAYQATVCARAQAPPPSDQMKCLIRLKLPTLHGPTPL 1629
Qy 601 YRLGAVQNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCWVIVGR 660
Db 1630 YRLGAVTNEVTLTHPTVKYIATCMQADLEVTSTWLVGGVLAALAAAYCLATGCVCIIGR 1689
Qy 661 VVLSGKAIPITDREVLYREFDEMEEC 686
Db 1690 LHVNQRAVAVAPDKREVLYEAFDEMEEC 1715

RESULT 11
GNVJ08
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J8)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A40250; PQ0397; PQ0559
R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Iizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1;
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
```

A:Molecule type: genomic RNA  
A:Residues: 2678-2754 <CHA>  
A:Cross-references: DDBJ:D10134  
R:Kato, N.; Ootsubayama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno  
Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
A:Title: Distribution of plural HCV types in Japan.  
A:Reference number: PQ0554; MUID:92068204; PMID:1720309  
A:Accession: PQ0559  
A:Molecule type: mRNA  
A:Residues: 2678-2729 <KAT>  
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PID:BAA01418.1; PID:g221524  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein C; status predicted <CPC>  
F:115-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEE>  
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1011-1619/Product: hepatitis virus #status predicted <NS3>  
F:1234-1241/Product: nucleotide-binding motif A (P-loop)  
F:1316-1321/Product: nucleotide-binding motif B  
F:1320-1323/Product: DEXH motif  
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4a>  
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>  
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 82.3%; Score 2979; DB 1; Length 3033;  
Best Local Similarity 79.7%; Pred. No. 7.3e-197;  
Matches 547; Conservative 62; Mismatches 17; Indels 0; Gaps 0;

QY 1 MAPITAYAOQTRGLGCIITSLTRDKNQVEGEVOIVSTAAOFLATCINGVMTVYHGA 60  
DB 1030 LAPITAYTOQTRGLGAIIVSLTRDKNEAGQVQLSSVYTOFLGTSIGVLWTVYHGA 1089

QY 61 GRTIASPKGPVQMTNVNDDLVGWPAPQASRLTCTCGSDLYLVTRHADVIPVRR 120  
DB 1090 GNTLAGPKGPVQMTVTSAGDLVGPSPCTGSLDCTCGAVDLVLTNADVIPVRR 1149

QY 121 GDSRGILLSRPRIYSLKGGSGGFLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1150 DRRGALLSRPLSTLKGSGGVPVLCRSHAGVGLFRAAVCARGVAKSIDIPVESLDVAT 1209

QY 181 RSPVFTDNSPPVQSFQVAHLHAPGSGKSTKVPAAVAAQGVKVLNPSVAATLGFG 240  
DB 1210 RTFSFSDNSTPPVQSFQVGYLHAPGSGKSTKVPAAVAAQGVKVLNPSVAATLGFG 1269

QY 241 AYMSKAHGIDPNRTGVTITTSPTITSTYTYGFLADGGCGGAYDIIICDECHSTDATS 300  
DB 1270 AYMSKAHGIDPNRTGVTITTSPTITSTYTYGFLADGGCGGAYDIIICDECHSTDATS 1329

QY 301 ILGIGTVLDOAETAGARLVVLTATPPGVSVTVPHPNIEEVALSTTGIPYGAIPLEVI 360  
DB 1330 ILGIGTVLDOAETAGARLVVLTATPPGVSVTVPHPNIEEVALSTTGIPYGAIPLEVI 1389

QY 361 KGRHLIFCHSKKCDLAALVAGLVNAYVYRGDLVSVIPPIGDVVVATDALMTGYT 420  
DB 1390 KGRHLIFCHSKKCDLAALVAGLVNAYVYRGDLVSVIPPIGDVVVATDALMTGYT 1449

QY 421 GDFSDVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRTOGRGTPGPIYRFVAPG 480  
DB 1450 GDFSDVIDCNTCTVTQTVDFSLDPTFTIETITLPQDAVSRTOGRGTPGPIYRFVAPG 1509

QY 481 ERPSGMFSDSVLCECYDAGCAWYELTTPAETTVLRAYNTPGLPVCQDHLFEWEGVTGL 540  
DB 1510 ERPSGMFSDSVLCECYDAGCAWYELTTPAETTVLRAYNTPGLPVCQDHLFEWEGVTGL 1569

QY 541 THIDAHFLSQTQKSGENLPYLVAQATVCARAKAPPPSDOMWKKLRLKPTLHGPTPL 600  
DB 1570 THIDAHFLSQTQKSGENLPYLVAQATVCARAKAPPPSDOMWKKLRLKPTLHGPTPL 1629

QY 601 YRLGAVQNEITLTHPVTKYIMTMSADLEVVTTSTWLVGGVLAALAAAYCLSTGCVCVIVGR 660

DB 1630 YRLGAVTNEVTLTHPVTKYIATCMQADLEIMTSWVLAAGVLAAYAAVCLATGCTSIIGR 1689

QY 661 VVLSCKPAIIPDREVLVYREFDEMEEC 686  
DB 1690 LHLNDRVVVAPDKELIYEAFFDEMEEC 1715

RESULT 12  
PC2219  
polypeptide - hepatitis C virus (type 5a) (fragments)  
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A protei  
C:Species: hepatitis C virus  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
A:Accession: PC2219  
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the he  
A:Reference number: PC2219; MUID:94338342; PMID:7520237  
A:Accession: PC2219  
A:Molecule type: mRNA  
A:Residues: 1-876 <STU>  
A:Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579  
A:Experimental source: serum  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: glycoprotein  
F:1-191/Product: core #status predicted <COE>  
F:58-78/Region: variable  
F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F:248-338/Region: E2  
F:339-411/Region: NS1 (amino end)  
F:412-783/Product: NS3 #status predicted <NSR>  
F:784-837/Product: NS4A #status predicted <NSA>  
F:838-876/Product: NS4B #status predicted <NSB>  
F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.1%; Score 1993; DB 2; Length 876;  
Best Local Similarity 85.5%; Pred. No. 1.6e-129;  
Matches 365; Conservative 31; Mismatches 31; Indels 0; Gaps 0;

QY 260 ITTGSPITSTYTKGFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDOAETAGARLV 319  
DB 411 ITTGASITSTYTKGFLADGGCGGAYDIIICDECHSQDATTILGIGTVLDOAETAGARLV 470

QY 320 VLATATPPGVSVTVPHPNIEEVALSTTGIPYGAIPLEVIKGRHLIFCHSKKCDLA 379  
DB 471 VLATATPPGVSVTVPHPNIEEVALPQGEVFFYGRAIPLEVIKGRHLIFCHSKKCDLA 530

QY 380 AKLVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYTGFSDVIDCNTCTVTQTVDF 439  
DB 531 KQLTSLGVNAYVYRGDLVAVIPTAGDVVVVCTDALMTGTFDSDVIDCNSAVTQTVDF 590

QY 440 SLDPPTTETITLPQDAVSRTOGRGTPGPIYRFVAPGERSGMFSDSVLCECYDAG 499  
DB 591 SLDPPTTETITLPQDAVSRTOGRGTPGPIYRFVAPGERSGMFSDSVLCECYDAG 650

QY 500 CAVVELTTPAETTVLRAYNTPGLPVCQDHLFEWEGVFTGLTHIDAHFLSQTQKSGENLP 559  
DB 651 CAVYDLTTPAETTVLRAYNTPGLPVCQDHLFEWEGVFTGLTHIDAHFLSQTQKSGENFP 710

QY 560 YLVAYQATVCARAKAPPPSDOMWKKLRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 619  
DB 711 YLVAYQATVCARAKAPPPSDOMWKKLRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 770

QY 620 IMTMSADLEVVTTSTWLVGGVLAALAAAYCLSTGCVCVIVGRVLSGKPAIIPDREVLRE 679  
DB 771 IMACMSADLEVVTTSTWLVGGVLAALAAAYCLTGVSAIVGRIILSGKPAIIPDREVLQQ 830

QY 680 FDEMEEC 686  
DB 831 FDEMEEC 837

RESULT 13  
S68016  
ATPase/RNA helicase - hepatitis C virus (fragment)  
C;Species: hepatitis C virus  
C;Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: S68016  
R;Jin, L.; Peterson, D.L.  
Arch. Biochem. Biophys. 323, 47-53, 1995  
A;Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
A;Reference number: S68016; MUID:96019946; PMID:7487072  
A;Accession: S68016  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-386 <JIN>  
A;Cross-references: UNIPROT:Q04045  
C;Superfamily: hepatitis C virus genome polypeptide  
C;Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polypeptide  
F;24-31/Region: nucleotide-binding motif A (P-loop)  
F;86-91/Region: nucleotide-binding motif B  
F;90-93/Region: DEXH motif

Query Match 54.5%; Score 1972.5; DB 2; Length 386;  
Best Local Similarity 92.4%; Pred. No. 1.4e-128;  
Matches 375; Conservative 3; Mismatches 7; Indels 21; Gaps 2;

Qy 183 PVFTDN--SSPPVPSQFVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 241  
Db 1 PVFTDNSSPPVPSQFVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATLGFGA 60  
Qy 242 YMSKAHGIDNIRTVGRTITGSPITVYTKGLADGCGSGGAYDIIICDECHSTDATSI 301  
Db 61 YMSKAHGVD-----YCKFLADGCGSGGAYDIIICDECHSTDATSI 100  
Qy 302 LGTGTVDQAEATAGARLVATATATPGSVTVPHNIEEVALSTGEIPFYKALPLEVIK 361  
Db 101 LGTGTVDQAEATAGARLVATATATPGSVTVPHNIEEVALSTGEIPFYKALPLEBAIK 160  
Qy 362 GGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYTG 421  
Db 161 GGRHLIFCHSKKKKDELAATKLVALGINAVAYRGLDVSVPSSGDDVVVATDALMTGFSG 220  
Qy 422 DFDSVIDCNTCVTQTVDFSLDPTFTIETILPDQAVSRTRGRGTGKPGIYRFVAPGE 481  
Db 221 DFDSVIDCNTCVTQTVDFSLDPTFTIETILPDQAVSRTRGRGTGKPGIYRFVAPGE 280  
Qy 482 RPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLT 541  
Db 281 RPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLT 340  
Qy 542 HIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLI 587  
Db 341 HIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLI 386

RESULT 14  
PS0326  
polypeptide - hepatitis C virus (isolate Fla) (fragments)  
C;Species: hepatitis C virus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: PS0326  
R;Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.  
Gene 105, 167-172, 1991  
A;Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype  
A;Reference number: PS0326; MUID:92039028; PMID:1718820  
A;Accession: PS0326  
A;Molecule type: Genomic RNA  
A;Residues: 1-492 <LIJ>  
A;Cross-references: UNIPROT:Q91F5; UNIPROT:Q36579; UNIPROT:Q036610; UNIPROT:Q03463; UNIPROT:Q0320  
A;Note: this sequence corresponds to nonstructural protein NS3 region  
A;Note: translation of the nucleotide sequence is not complete  
C;Superfamily: hepatitis C virus genome polypeptide

C;Keywords: polypeptide

Query Match 38.4%; Score 1388; DB 2; Length 492;  
Best Local Similarity 98.1%; Pred. No. 4.1e-88;  
Matches 257; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 425 SVIDCNTCVTQTVDFSLDPTFTIETILPDQAVSRTRGRGTGKPGIYRFVAPGERPS 484  
Db 1 SVIDCNTCVTQTVDFSLDPTFTIETILPDQAVSRTRGRGTGKPGIYRFVAPGERPS 60  
Qy 485 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLTHID 544  
Db 61 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLTHID 120  
Qy 545 AHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLIIRLKLPTLHGPTPLLLYRLG 604  
Db 121 AHFLSQTQSGENLPYLVAQATVCARAQAPPPSWDMWKLIIRLKLPTLHGPTPLLLYRLG 180  
Qy 605 AVQNEITLTHPVTKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCVIVGVVLS 664  
Db 181 AVQNEITLTHPVTKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCVIVGVVLS 240  
Qy 665 GKPAIIPDREVLYREFDEMEEC 686  
Db 241 GKPAIIPDREVLYREFDEMEEC 262

## RESULT 15

JQ1366

polypeptide - hepatitis C virus (French isolate) (fragments)

C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C;Accession: JQ1366

R;Kremers, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implication

A;Reference number: JQ1366; MUID:92013977; PMID:1655961

A;Accession: JQ1366

A;Molecule type: Genomic RNA

A;Residues: 1-716 &lt;KRE&gt;

A;Cross-references: UNIPROT:Q9PX22

C;Superfamily: hepatitis C virus genome polypeptide

C;Keywords: Glycoprotein; polypeptide

F;84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #stat

Query Match 38.2%; Score 1383; DB 2; Length 716;

Best Local Similarity 51.8%; Pred. No. 1.5e-87;

Matches 318; Conservative 43; Mismatches 103; Indels 150; Gaps 21;

Qy 132 PISYKSGSGGLLCPAGHAGIFRAAVCTRGVAKAVDFIPVENLETMRSPVFTDNSSP 191

Db 138 PISYANGT--GPEHRP-----YCWHPKPGIYPAQ---TVCGPVYCFPTSP 180

Qy 192 PVYPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATLGL--FG-----AYM 243

Db 181 VVVGTTNK---LGAPTYN-----WGNDDTVFLN--NTRPPLGNWFGCTWNSGGF 227

Qy 244 SKAHGIDPNIRTVGRTITGSP-----TSTYSGKFLADGCGSGGAYDIIICDECHSTD 297

Db 228 TKVCGAPCVIGGAGNNTLYCPDCKFKHPEATYSR-----CGSGPW---ITPRC----- 274

Qy 298 ATSLG-----IGTV-----LDQATAGARL 318

Db 275 ---LVGYPRLMHPCTVNTLTKVRYMGVGVHRLQVACNWTGRGNCNDRDRSELSP 331

Qy 319 WLAT-----ATPGSVTVPHNIEEVALSTGEIPFYKALPLEVIKGRHLIFCHSKKK 374

Db 332 LLSTQWQVLPSCFTTLP-----AL--TGLHLHQNIVDQVLYG----- 371

Qy 375 CDELAALKVALGINAVAYRGLDVSVPPIGDDVVVATDALMTGYTGFDSVIDC--NTC 432

Db 372 -----VGSISVWAKWEVIL-----LFLLLADA-----RVCCLWNTC 406

Qy	433	VTQTVDFSLDPTFTIETITLPQDAVSRTORRGRTGRGKPGIYRFVAPGERPSCMFDSVL	492
Db	407	VTQTVDFSLDPTFTIETITLPQDAVSRTORRGRTGRGKPGIYRFVAPGERPSCMFDSVL	466
Qy	493	CECYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSOTK	552
Db	467	CECYDAGCAWYELTPAETTVRLRAYNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSOTK	526
Qy	553	QSGENLPYLVAQATVCARAQAPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITL	612
Db	527	QSGENLPYLVAQATVCARAQAPPSWDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEVTL	586
Qy	613	THPVTKYIMTCSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRVVLSCKPAIIPD	672
Db	587	THPITKYIMTCSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRVILSKPAIIPD	646
Qy	673	REVLRYREFDEMEEC	686
Db	647	REVLRYREFDEMEEC	660

Search completed: November 7, 2005, 20:10:53  
Job time : 17.1714 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 87.9151 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-2

Perfect score: 3619

Sequence: 1 MAPITAYAQTRGLGCIIT.....PAIIPREVLYREFDEMEEC 686

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3602	99.5	2436	2 Q81756	Q81756 hepatitis c
2	3602	99.5	3011	1 POLG_HCV1	P26664 h genome po
3	3598	99.4	3011	2 Q1FPE5	Q1FPE5 hepatitis c
4	3574	98.8	2908	2 Q61X04	Q61X04 hepatitis c
5	3560	98.4	3011	2 Q36579	Q36579 hepatitis c
6	3553	98.2	3011	2 Q36608	Q36608 hepatitis c
7	3553	98.2	3015	2 Q9PW9	Q9PW9 hepatitis c
8	3549	98.1	3011	2 Q9EL58	Q9EL58 hepatitis c
9	3549	98.1	3011	2 Q36609	Q36609 hepatitis c
10	3542	97.9	3011	2 Q36610	Q36610 hepatitis c
11	3542	97.9	3011	2 Q3463	Q3463 hepatitis c
12	3542	97.9	3011	2 Q9DIT6	Q9DIT6 hepatitis c
13	3540	97.8	3011	2 Q9DIT6	P27958 h genome po
14	3513	97.1	3011	1 POLG_HCVH	Q9DTE2 hepatitis c
15	3436	94.9	3010	2 Q9DTE2	Q9DIY2 hepatitis c
16	3436	94.9	3010	2 Q9DIY1	Q9DIY1 hepatitis c
17	3433	94.9	3010	2 Q9J3G2	P26663 h genome po
18	3432	94.8	3010	1 POLG_HCVBK	Q9DTE7 hepatitis c
19	3426	94.7	3010	2 Q9DTE7	Q9DIY2 hepatitis c
20	3426	94.7	3010	2 Q9DTE7	Q9DIY2 hepatitis c
21	3426	94.7	3010	2 Q9DTE7	Q9DIY2 hepatitis c
22	3425	94.6	3010	2 Q9PAU2	Q9PAU2 hepatitis c
23	3425	94.6	3010	2 Q68788	Q68788 hepatitis c
24	3425	94.6	3010	2 Q9J3H6	Q9J3H6 hepatitis c
25	3425	94.6	3013	2 Q636P5	Q636P5 hepatitis c
26	3424	94.6	3010	2 Q9J3H0	Q9J3H0 hepatitis c
27	3424	94.6	3010	2 Q9J3H8	Q9J3H8 hepatitis c
28	3422	94.6	3013	2 Q9QNC0	Q9QNC0 hepatitis c
29	3421	94.5	3010	2 Q9QIX1	Q9QIX1 hepatitis c
30	3421	94.5	3010	2 Q9QIY9	Q9QIY9 hepatitis c
31	3421	94.5	3010	2 Q9QIZ0	Q9QIZ0 hepatitis c

#### RESULT 1

ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996	(TrEMBLrel. 01, Created)		
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Polyprotein (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; asRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus			
OX	NCBI_taxid=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: M32084; AAA45677.1; -			
DR	PIR: PS0326; PS0326.			
DR	PIR: PS0327; PS0327.			
DR	PIR: PS0328; PS0328.			
DR	HSSP: P27958; 1A1V.			
DR	GO: GO:0016021; C:integral to membrane; IEA.			
DR	GO: GO:0019038; C:viral capsid; IEA.			
DR	GO: GO:0019031; C:viral envelope; IEA.			
DR	GO: GO:0005524; F:ATP binding; IEA.			
DR	GO: GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO: GO:0003723; F:RNA binding; IEA.			
DR	GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO: GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO: GO:0005198; F:structural molecule activity; IEA.			
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO: GO:0006350; P:transcription; IEA.			
DR	GO: GO:0019079; P:viral genome replication; IEA.			
DR	GO: GO:0019087; P:viral transformation; IEA.			
DR	InterPro: IPR000345; CytC_heme_BS.			
DR	InterPro: IPR001410; DEAD.			
DR	InterPro: IPR011545; HCV NS1.			
DR	InterPro: IPR002531; HCV NS1.			
DR	InterPro: IPR000745; HCV NS4a.			
DR	InterPro: IPR001490; HCV NS4b.			
DR	InterPro: IPR002868; HCV NS4c.			
DR	InterPro: IPR002166; HCV NS4d.			
DR	InterPro: IPR001550; Helicase_C.			
DR	InterPro: IPR004109; Peptidase_S29.			
DR	InterPro: IPR009003; Pept_Ser_Cys.			
DR	InterPro: IPR002518; Pept_U39_HCV_NS2.			
DR	InterPro: IPR007895; RNA_pol_DS_P5.			
DR	InterPro: IPR007094; RNA_pol_Fsvir.			
DR	Pfam: PF01560; HCV NS1; 1.			
DR	Pfam: PF01538; HCV NS2; 1.			
DR	Pfam: PF02907; HCV NS3; 1.			
DR	Pfam: PF01006; HCV NS4a; 1.			
DR	Pfam: PF01001; HCV NS4b; 1.			
DR	Pfam: PF01506; HCV NS5a; 1.			

#### ALIGNMENTS

32	3421	94.5	3013	2	Q9QIY0	Q9QIY0 hepatitis c
33	3421	94.5	3014	2	Q6GYR8	Q6GYR8 hepatitis c
34	3420	94.5	3010	2	Q9DTE5	Q9DTE5 hepatitis c
35	3420	94.5	3010	2	Q9J3G8	Q9J3G8 hepatitis c
36	3420	94.5	3010	2	Q9J3G9	Q9J3G9 hepatitis c
37	3420	94.5	3013	2	Q9QIX9	Q9QIX9 hepatitis c
38	3419	94.5	3010	2	Q6GYR9	Q6GYR9 hepatitis c
39	3419	94.5	3010	2	Q9DTE0	Q9DTE0 hepatitis c
40	3419	94.5	3010	2	Q9QIY5	Q9QIY5 hepatitis c
41	3419	94.5	3010	2	Q9QIY6	Q9QIY6 hepatitis c
42	3419	94.5	3015	2	Q9WPH5	Q9WPH5 hepatitis c
43	3418	94.4	3010	2	Q9DTE6	Q9DTE6 hepatitis c
44	3418	94.4	3010	2	Q9J3F9	Q9J3F9 hepatitis c
45	3417	94.4	3010	1	POLG_HCVJT	Q00269 h genome po



KW Envelope protein; Glycoprotein; Helicase; Hydrolase;  
KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
KW Serine protease; Transferase; Transmembrane.  
FT INIT\_MET 1 1  
FT CHAIN 1 115  
FT CHAIN 116 191  
FT CHAIN 192 383  
FT CHAIN 384 729  
FT CHAIN 730 1006  
FT CHAIN 1007 1615  
FT CHAIN 1616 1862  
FT CHAIN 1863 2013  
FT CHAIN 2014 3011  
FT TRANSMEM 347 369  
FT ACT\_SITE 1083 1083  
FT ACT\_SITE 1107 1107  
FT ACT\_SITE 1165 1165  
FT NP\_BIND 1230 1237  
FT SITE 1316 1319  
FT CARBOHYD 196 196  
FT CARBOHYD 209 209  
FT CARBOHYD 234 234  
FT CARBOHYD 305 305  
FT CARBOHYD 417 417  
FT CARBOHYD 423 423  
FT CARBOHYD 430 430  
FT CARBOHYD 448 448  
FT CARBOHYD 476 476  
FT CARBOHYD 532 532  
FT CARBOHYD 540 540  
FT CARBOHYD 556 556  
FT CARBOHYD 576 576  
FT CARBOHYD 623 623  
FT CARBOHYD 645 645  
FT CARBOHYD 2041 2041  
FT CARBOHYD 2077 2077  
FT CARBOHYD 2240 2240  
FT CARBOHYD 2364 2364  
FT CARBOHYD 2789 2789  
FT STRAND 1224 1226  
FT TURN 1232 1234  
FT TURN 1236 1238  
FT HELIX 1239 1246  
FT TURN 1247 1248  
FT STRAND 1251 1255  
FT STRAND 1291 1295  
FT HELIX 1296 1301  
FT TURN 1302 1303  
FT TURN 1312 1315  
FT TURN 1316 1319  
FT HELIX 1323 1335  
FT TURN 1336 1336  
FT HELIX 1337 1339  
FT TURN 1340 1340  
FT STRAND 1343 1346  
FT TURN 1352 1353  
FT STRAND 1363 1366  
FT STRAND 1373 1373  
FT TURN 1376 1377  
FT STRAND 1380 1380  
FT STRAND 1382 1385  
FT STRAND 1390 1393  
FT HELIX 1397 1408  
FT TURN 1409 1411  
FT STRAND 1414 1417  
FT TURN 1430 1431  
FT STRAND 1433 1436  
FT TURN 1438 1439  
FT TURN 1442 1444  
FT STRAND 1450 1453  
FT STRAND 1456 1463  
FT STRAND 1471 1478

FT HELIX 1481 1488  
FT TURN 1489 1490  
FT STRAND 1498 1501  
FT HELIX 1514 1526  
FT TURN 1527 1527  
FT HELIX 1527 1544  
FT TURN 1532 1547  
FT HELIX 1546 1547  
FT TURN 1555 1563  
FT TURN 1564 1564  
FT HELIX 1570 1578  
FT TURN 1579 1580  
FT HELIX 1584 1597  
FT TURN 1598 1598  
FT HELIX 1606 1611  
FT TURN 1614 1618  
FT STRAND 1622 1623  
FT STRAND 1627 1627  
FT STRAND 1635 1636  
FT HELIX 1640 1652  
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;  
Query Match 99.5%; Score 3602; DB 1; Length 3011;  
Best Local Similarity 99.6%; Pred. No. 8.3e-244;  
Matches 683; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MAPTAYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 60  
DB 1026 LAPITAYAAQTRGLLGCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTYYHGA 1085  
QY 61 GTRTIASPKGPVIQMYTNVDQDLVGMWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRR 120  
DB 1086 GTRTIASPKGPVIQMYTNVDQDLVGMWPAQPSRSITPCTCGSSDLYLVTRHADVIPVRRR 1145  
QY 121 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
DB 1146 GDSRGSLLSPRPISYLGKSGGPLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
QY 181 RSPVFTDNSSPPVPPQSPQVAHLHAPTQSGSKSTKVPAAYAAQGYKVLVLPNSVAATLGF 240  
DB 1206 RSPVFTDNSSPPVPPQSPQVAHLHAPTQSGSKSTKVPAAYAAQGYKVLVLPNSVAATLGF 1265  
QY 241 AYMSKAHGIDPNIRGTVRTITTTGSPITVTYTGKFLADGCGSGGAYDIIICDECHSTDATS 300  
DB 1266 AYMSKAHGIDPNIRGTVRTITTTGSPITVTYTGKFLADGCGSGGAYDIIICDECHSTDATS 1325  
QY 301 ILGIGTVLDQAEATAGARLVWLATATPPGSGVTVPHPNIEEVALSTTGEIPFYGKAIPLEVI 360  
DB 1326 ILGIGTVLDQAEATAGARLVWLATATPPGSGVTVPHPNIEEVALSTTGEIPFYGKAIPLEVI 1385  
QY 361 KGGRHILFCHSKKKCDELAALKVALGINAVAYYRGLDVSVIPPTGDDVVVATDALMTGYT 420  
DB 1386 KGGRHILFCHSKKKCDELAALKVALGINAVAYYRGLDVSVIPPTGDDVVVATDALMTGYT 1445  
QY 421 GDFDSVIDCNTCVTQTVDFSLDPTETITILPDQAVSRTORRGTRGKPGIYRFVAPG 480  
DB 1446 GDFDSVIDCNTCVTQTVDFSLDPTETITILPDQAVSRTORRGTRGKPGIYRFVAPG 1505  
QY 481 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPCQDHFLEWEGVFTGL 540  
DB 1506 ERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPCQDHFLEWEGVFTGL 1565  
QY 541 THIDAHFLSQKSGENIPYLVAQATVCARAQAPPPSWDQWKKLIRLKPTLHGPTPLL 600  
DB 1566 THIDAHFLSQKSGENIPYLVAQATVCARAQAPPPSWDQWKKLIRLKPTLHGPTPLL 1625  
QY 601 YRLGAVQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSGCVVIVGR 660  
DB 1626 YRLGAVQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSGCVVIVGR 1685  
QY 661 VVLSGKPAIIPDREVLXREFDEMEEC 686  
DB 1686 VVLSGKPAIIPDREVLXREFDEMEEC 1711

```
RESULT 3
Q91F55 PRELIMINARY; PRT; 3011 AA.
AC Q91F55;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
RT sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
DR EMBL: AF271632; AAF81759.1; -.
DR PIR: A4150; A44150.
DR PIR: Q0804; Q0804.
DR PIR: PS0326; PS0326.
DR PIR: PS0327; PS0327.
DR PIR: PS0328; PS0328.
DR HSP: O8JYS1; ICWX.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003668; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019079; P: viral genome replication; IEA.
DR GO: GO:0019087; P: viral transformation; IEA.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH N.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_S29.
DR InterPro: IPR090003; Pept_Ser_Cys.
DR InterPro: IPR002518; Pept_U39_HCV_NS2.
DR InterPro: IPR007095; RNA_pol_Ds_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;
```

```
Query Match 99.4%; Score 3598; DB 2; Length 3011;
Best Local Similarity 99.4%; Pred. No. 1.6e-243; Indels 0; Gaps 0;
Matches 682; Conservative 2; Mismatches 2;

Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATFCINGVCMWTVYHGA 60
Db 1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVGEVQIVSTAAQTFLATFCINGVCMWTVYHGA 1085
Qy 61 GTRTIASPKGPVIQMTYNDQDLVGHAPAGOGSSLSLTCTCGSSDLYLVTSHADVIPVRRR 120
Db 1086 GTRTIASPKGPVIQMTYNDQDLVGHAPAGOGSSLSLTCTCGSSDLYLVTSHADVIPVRRR 1145
Qy 121 GDSRGSLLSPRISYLYKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLSTTM 180
Db 1146 GDSRGSLLSPRISYLYKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLSTTM 1205
Qy 181 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 240
Db 1206 RSPVFTDNSSPPVVPVQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLGFG 1265
Qy 241 AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 300
Db 1266 AYMSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSTDATS 1325
Qy 301 ILGIGTVLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGPIPVGKAIPLEVI 360
Db 1326 ILGIGTVLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGPIPVGKAIPLEVI 1385
Qy 361 KGRHLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVATDALMTGYT 420
Db 1386 KGRHLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVATDALMTGYT 1445
Qy 421 GDFDSVIDCNTCTQTVDFSLDFTFETITLTPQDAVSRTQRRGTRGKPGIYRFVAPG 480
Db 1446 GDFDSVIDCNTCTQTVDFSLDFTFETITLTPQDAVSRTQRRGTRGKPGIYRFVAPG 1505
Qy 481 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 540
Db 1506 ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 1565
Qy 541 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMMKCLIRLKTPLHGPPTLL 600
Db 1566 THIDAHFLSQTKSGENLPYLVAQATVCARAQAPPSQDMMKCLIRLKTPLHGPPTLL 1625
Qy 601 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGVLAALAAAYCISTGCVNLVGR 660
Db 1626 YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGVLAALAAAYCISTGCVNLVGR 1685
Qy 661 VLSGKPAIIPDREVLYRPEDEMEEC 686
Db 1686 VLSGKPAIIPDREVLYRPEDEMEEC 1711

RESULT 4
Q6IX04 PRELIMINARY; PRT; 2908 AA.
AC Q6IX04; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1a;
RA Brann T.W., Kottlil S., Polis M., Imamichi T.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY615798; AAY44836.1; -.
DR HSHP; P27958; 1A1V.
DR
```



DR GO: 0016021; C: integral to membrane; IEA.  
 DR GO: 0019028; C: viral capsid; IEA.  
 DR GO: 0019031; C: viral envelope; IEA.  
 DR GO: 0005524; F: ATP binding; IEA.  
 DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
 DR GO: 0016787; F: hydrolase activity; IEA.  
 DR GO: 0003723; F: RNA binding; IEA.  
 DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 0008236; F: serine-type peptidase activity; IEA.  
 DR GO: 0005198; F: structural molecule activity; IEA.  
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: 0006350; P: transcription; IEA.  
 DR GO: 0019079; P: viral genome replication; IEA.  
 DR GO: 0019087; P: viral transformation; IEA.  
 DR InterPro: IPR00345; CytC\_heme\_BS.  
 DR InterPro: IPR011410; DEAD.  
 DR InterPro: IPR011545; DEAD/DEAH\_N.  
 DR InterPro: IPR02522; HCV\_capsid.  
 DR InterPro: IPR02521; HCV\_core.  
 DR InterPro: IPR02519; HCV\_env.  
 DR InterPro: IPR02531; HCV\_NS1.  
 DR InterPro: IPR00745; HCV\_NS4a.  
 DR InterPro: IPR001490; HCV\_NS4b.  
 DR InterPro: IPR02868; HCV\_NS5a.  
 DR InterPro: IPR002166; HCV\_RdRP.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV\_capsid; 1.  
 DR Pfam: PF01542; HCV\_core; 1.  
 DR Pfam: PF01539; HCV\_env; 1.  
 DR Pfam: PF01560; HCV\_NS1; 1.  
 DR Pfam: PF01538; HCV\_NS2; 1.  
 DR Pfam: PF01006; HCV\_NS4a; 1.  
 DR Pfam: PF01001; HCV\_NS4b; 1.  
 DR Pfam: PF01506; HCV\_NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.  
 FT NON\_TER 2908 2908  
 SQ SEQUENCE 2908 AA; 315737 MW; BFS44BC591498A4F CRC64;  
  
 Query Match 98.8%; Score 3574; DB 2; Length 2908;  
 Best Local Similarity 97.7%; Pred. No. 7.4e-242;  
 Matches 670; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
  
 Qy 1 MAPITAAQTRGLGCIITSLTRDKKNQVEGEIVSTAAQTFLATCINGCVTVYHGA 60  
 Db 1026 LAPITAAQTRGLGCIITSLTRDKKNQVEGEIVSTAAQTFLATCINGCVTVYHGA 1085  
  
 Qy 61 GTRIASPGKGVIMYNTVDQDVGWPAQGSRLTPTCTGSSDLYLVTRHADVIPVRRR 120  
 Db 1086 GTRIASPGKGVIMYNTVDQDVGWPAQGSRLTPTCTGSSDLYLVTRHADVIPVRRR 1145  
  
 Qy 121 GDSRGLSPRPISYLVKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
 Db 1146 GDSRGLSPRPISYLVKSGSGGLPCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
  
 Qy 181 RSPVFTDNSSPPVQSFQVAHLHAPTCGSKSTKVPAAAGQKVLVLPNSVAATLGF 240  
 Db 1206 RSPVFTDNSSPPVQSFQVAHLHAPTCGSKSTKVPAAAGQKVLVLPNSVAATLGF 1265  
  
 Qy 241 AYMSKAHGIDNIRTVRTITGSPITYSTYKFLADGCGSGGAYDIICDECHSTDATS 300  
 Db 1266 AYMSKAHGVDNIRTVRTITGSPITYSTYKFLADGCGSGGAYDIICDECHSTDATS 1325  
  
 Qy 301 ILGIGTLDQAETAGARLVLAATATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEVI 360

Db 1326 ILGIGTLDQAETAGARLVLAATATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEAI 1395  
 Qy 361 KGGRHILFCHSKKCDLAALVALGINAVAYYRGLDVSVIPPIGDVVVATDALMTGYT 420  
 Db 1386 KGGRHILFCHSKKCDLAALVALGINAVAYYRGLDVSVIPPTSGDVVVVATDALMTGFT 1445  
 Qy 421 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLQDAVSRQRTGRGKGIYRFVAPG 480  
 Db 1446 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLQDAVSRQRTGRGKGIYRFVAPG 1505  
 Qy 481 ERPSGMFDSVLCCEVDAGCAWYELPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 540  
 Db 1506 ERPSGMFDSVLCCEVDAGCAWYELPAETTVRLRAYMNTPGLPVCDHLEFEGVFTGL 1565  
 Qy 541 THIDAHFLSQKSGENLPYLVAIQATVCARAQPPSDQMKCLIRLKLPTLHGPTPL 600  
 Db 1566 THIDAHFLSQKSGENLPYLVAIQATVCARAQPPSDQMKCLIRLKLPTLHGPTPL 1625  
 Qy 601 YRLGAVONEITLTHPVTKYIMTCHSADLEVTSTWLVGVLAALAAAYCLSTGCWIVGR 660  
 Db 1626 YRLGAVONEITLTHPVTKYIMTCHSADLEVTSTWLVGVLAALAAAYCLSTGCWIIGR 1685  
 Qy 661 VVLSGKPAIIPDREVLYRBEFDEMEEC 686  
 Db 1686 IVLSGKPAVIPDREVLYRBEFDEMEEC 1711  
  
 RESULT 5  
 O36579 PRELIMINARY; PRT; 3011 AA.  
 AC O36579;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Polypotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 CX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;  
 RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,  
 Rice C.M.;  
 RT "Transmission of hepatitis C by intrahepatic inoculation with  
 transcribed RNA.";  
 RL Science 277:570-574(1997).  
 DR EMBL; AF009606; AA866324.1; -.  
 DR PIR; A44150; A44150.  
 DR PIR; PQ0804; PQ0804.  
 DR PIR; PS0326; PS0326.  
 DR PIR; PS0327; PS0327.  
 DR PIR; PS0328; PS0328.  
 DR PDB; 1N1L; X-ray; A/B=1017-1214.  
 DR GO: 0019028; C: viral capsid; IEA.  
 DR GO: 0019031; C: viral envelope; IEA.  
 DR GO: 0005524; F: ATP binding; IEA.  
 DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
 DR GO: 0003723; F: RNA binding; IEA.  
 DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 0008236; F: serine-type peptidase activity; IEA.  
 DR GO: 0005198; F: structural molecule activity; IEA.  
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: 0006350; P: transcription; IEA.  
 DR GO: 0019079; P: viral genome replication; IEA.  
 DR GO: 0019087; P: viral transformation; IEA.  
 DR InterPro; IPR00345; CytC\_heme\_BS.  
 DR InterPro; IPR011410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH\_N.  
 DR InterPro; IPR02522; HCV\_capsid.  
 DR InterPro; IPR02521; HCV\_core.  
 DR InterPro; IPR02519; HCV\_env.  
 DR InterPro; IPR02531; HCV\_NS1.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR02868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.  
 FT NON\_TER 2908 2908  
 SQ SEQUENCE 2908 AA; 315737 MW; BFS44BC591498A4F CRC64;

DR	InterPro; IPR002531; HCV_NSI1.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR001490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRp.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR004109; Peptidase_S29.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF01543; HCV_capsid; 1.	
DR	Pfam; PF01542; HCV_core; 1.	
DR	Pfam; PF01539; HCV_env; 1.	
DR	Pfam; PF01538; HCV_NS1; 1.	
DR	Pfam; PF01538; HCV_NS2; 1.	
DR	Pfam; PF01538; HCV_NS3; 1.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF01006; HCV_NS4a; 1.	
DR	Pfam; PF01006; HCV_NS4b; 1.	
DR	Pfam; PF01006; HCV_NS4b; 1.	
DR	Pfam; PF00271; Helicase_C; 1.	
DR	Pfam; PF00998; Viral_RdRp; 1.	
DR	SMART; SM00487; DEXDC_1.	
DR	PROSITE; PS00140; CYTOCHROME C; UNKNOWN 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; Transmembrane.	
SQ	SEQUENCE 3011 AA; 327184 MW; E2E0EBE809C63C1B9 CRC64;	
Query Match 98.4%; Score 3560; DB 2: Length 3011;		
Best Local Similarity 97.7%; Pred. No. 7.5e-241;		
Matches 670; Conservative 9; Mismatches 7; Indels 0; Gaps 0;		
QY	1 MAPITAYAOQTRGLGCIITSLGRDNQVGEVQIVSTAAQTFLATCINGVCWTVYHGA 60	
DB	1026 LAPITAYAOQTRGLGCIITSLGRDNQVGEVQIVSTATQIFLATCINGVCWTVYHGA 1085	
QY	61 GTTITASPKGPVITOMYTNVDQDLVGMPPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 120	
DB	1086 GTRTITASPKGPVITOMYTNVDQDLVGMPPAPQGSRLTPTCTCGSSDLYLVTRHADVIPVRRR 1145	
QY	121 GDSRGSLLSPRPISYLGKSGGGPLLCAGHAGVIFRAAICTRGVAKAVDFIPVENLETTM 180	
DB	1146 GDSRGSLLSPRPISYLGKSGGGPLLCAGHAGVIFRAAICTRGVAKAVDFIPVENLETTM 1205	
QY	181 RSPVFTDSSPPVPPQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLIGFG 240	
DB	1206 RSPVFTDSSPPVPPQSFQVAHLHAPTGSCKSTKVPAAAYAAQGYKVLVLPNSVAATLIGFG 1265	
QY	241 AYMSKAHGDIPNIRTGVRTITTTGSPITYSTYVKFLADGGCGGGAYDIIICDECHSTDATS 300	
DB	1266 AYMSKAHGDIPNIRTGVRTITTTGSPITYSTYVKFLADGGCGGGAYDIIICDECHSTDATS 1325	
QY	301 ILGIGTVLDQAEATAGARLVLATATPGSVTVRHPNIEEVALSTTGEIPFYGKAIPLEVI 360	
DB	1326 ILGIGTVLDQAEATAGARLVLATATPGSVTVRHPNIEEVALSTTGEIPFYGKAIPLEVI 1385	
QY	361 KGGRHLLIFCHSKKKCDELAALVALGINAVAYRGLDVSVIPTSGDVVVVSTDALMTGTYT 420	
DB	1386 KGGRHLLIFCHSKKKCDELAALVALGINAVAYRGLDVSVIPTSGDVVVVSTDALMTGTYT 1445	
QY	421 GDFDSVTDNCVTQTQTVDFSLDPTFTTETTLPODAVSRQRRGTGRGPIYRFVAPG 480	
DB	1446 GDFDSVTDNCVTQTQTVDFSLDPTFTTETTLPODAVSRQRRGTGRGPIYRFVAPG 1505	
QY	481 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYMNTPLGVPCQDHLFEFEGVFTGL 540	
DB	1506 ERPSGMFSSVLCBCEYDAGCAWYELTPAETTVRLRAYMNTPLGVPCQDHLFEFEGVFTGL 1555	
QY	541 THIDAHFLSQTQSGENLPYLVAQYATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPLL 600	
DB	1566 THIDAHFLSQTQSGENFPYLVAQYATVCARAQAPPPSDQMMKCLIRLKLPTLHGPTPLL 1625	
QY	601 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCVVIGR 660	

Db	1626 YRLGAVQNEITLTHPVTKYIMTCMSADLEVVSTWLVGGVLAALAAAYCLSTGCVVIGR 1685	
QY	661 VVLSGKPAIPDPREVLYREFDEMEEC 686	
DB	1686 IVLSGKPAIPDPREVLYQEFDEMEEC 1711	
RESULT 6		
O36608 PRELIMINARY; PRT; 3011 AA.		
ID	O36608	
AC	O36608; (TrEMBLrel. 05, Created)	
DT	01-JAN-1998	
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Polyprotein.	
OS	Hepatitis C virus strain H77.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.	
OX	NCBI_TaxID=63746;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H77;	
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;	
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;	
RT	"Transcript from a single full-length cDNA clone of hepatitis C virus	
RT	are infectious when directly transfected into the liver of a	
RT	chimpanzee.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).	
DR	EMBL; AF011751; AAB67036.1; -.	
DR	PIR; A4150; A4150.	
DR	PIR; PQ0804; PQ0804.	
DR	PIR; PS0326; PS0326.	
DR	PIR; PS0327; PS0327.	
DR	PIR; PS0328; PS0328.	
DR	HSP; P27958; 1HEI.	
DR	GO; GO:0019028; C:viral capsid; IEA.	
DR	GO; GO:0019031; C:viral envelope; IEA.	
DR	GO; GO:0005524; F:ATP binding; IEA.	
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.	
DR	GO; GO:0003723; F:RNA binding; IEA.	
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.	
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.	
DR	GO; GO:0005198; F:structural molecule activity; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	GO; GO:0006350; P:transcription; IEA.	
DR	GO; GO:0019079; P:viral genome replication; IEA.	
DR	GO; GO:0019087; P:viral transformation; IEA.	
DR	InterPro; IPR000345; CytC_heme_BS.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR011545; DEAD/DEAH_N.	
DR	InterPro; IPR002522; HCV_capsid.	
DR	InterPro; IPR002521; HCV_core.	
DR	InterPro; IPR002519; HCV_env.	
DR	InterPro; IPR002531; HCV_NSI.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR001490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRp.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR004109; Peptidase_S29.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	
DR	Pfam; PF01543; HCV_capsid; 1.	
DR	Pfam; PF01542; HCV_core; 1.	
DR	Pfam; PF01539; HCV_env; 1.	
DR	Pfam; PF01538; HCV_NS1; 1.	
DR	Pfam; PF01538; HCV_NS2; 1.	
DR	Pfam; PF02907; HCV_NS3; 1.	
DR	Pfam; PF01006; HCV_NS4a; 1.	
DR	Pfam; PF01001; HCV_NS4b; 1.	

DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase C; 1.  
DR Pfam; PF00998; Viral RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 327114 MW; 0B75E6B81CB5C198 CRC64;  
  
Query Match 98.2%; Score 3553; DB 2; Length 3011;  
Best Local Similarity 97.5%; Pred. No. 2.3e-240;  
Matches 669; Conservative 9; Mismatches 8; Indels 0; Gaps 0;  
  
Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db :|||||  
1026 LAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTATQTFLATCINGVCWTVYHGA 1085  
Qy 61 GTRTIASPKGFVIMQYTNVDQDLVGVWPAQGSRLTCTCGSDLYLVTRHADVIPVRR 120  
Db :|||||  
1086 GTRTIASPKGFVIMQYTNVDQDLVGVWPAQGSRLTCTCGSDLYLVTRHADVIPVRR 1145  
Qy 121 GDSRGSLLSPRPISYLSKSGGGLLCPAGHAGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
Db :|||||  
1146 GDSRGSLLSPRPISYLSKSGGGLLCPAGHAGIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
Qy 181 RSPVFTDNSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGF 240  
Db :|||||  
1206 RSPVFTDNSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGF 1265  
Qy 241 AYMSKAHGDIPNRTGRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSDTATS 300  
Db :|||||  
1266 AYMSKAHGDIPNRTGRTITTSPTIYTYGKFLADGGCGGAYDIIICDECHSDTATS 1325  
Qy 301 ILGIGTVLDDQETAGARLVLATATPPGTVPHNPNEEVALSTTGIPFGKAIPLLEVI 360  
Db :|||||  
1326 ILGIGTVLDDQETAGARLVLATATPPGTVPHNPNEEVALSTTGIPFGKAIPLLEVI 1385  
Qy 361 KGRHLLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPIGDVVVATDALMTGYT 420  
Db :|||||  
1386 KGRHLLIFCHSKKCCDELAALVALGINAVAYRGLDVSVIPIGSDVVVSTDALMTGFT 1445  
Qy 421 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRQRRGRTGRGRIYRFPAG 480  
Db :|||||  
1446 GDFSDVDCNTCVTQTVDFSLDPTFTTETTLPODAVSRQRRGRTGRGRIYRFPAG 1505  
Qy 481 ERPSGMFDSVLCFCYDAGCAWYELTPAETTVLRAYMNTPLGVCDHLEFEGVFTGL 540  
Db :|||||  
1506 ERPSGMFDSVLCFCYDAGCAWYELTPAETTVLRAYMNTPLGVCDHLEFEGVFTGL 1565  
Qy 541 THIDAHFLSQTQKSGENLPYLVAQATVCARAQAPPSPSDQMWKCLIRLKPTLHGPTPL 600  
Db :|||||  
1566 THIDAHFLSQTQKSGENFPYLVAQATVCARAQAPPSPSDQMWKCLIRLKPTLHGPTPL 1625  
Qy 601 YRLGAVONEITLTPVTKYIMTCSADLEVTSTWLVGVLAALAYCYLSTGCWVIYGR 660  
Db :|||||  
1626 YRLGAVONEITLTPVTKYIMTCSADLEVTSTWLVGVLAALAYCYLSTGCWVIYGR 1685  
Qy 661 VVLGSKPAIIPDRVLVREFDEMEEC 686  
Db :|||||  
1686 IVLGGKPAIIPDRVLVREFDEMEEC 1711

RESULT 7

Q9PWU9 PRELIMINARY; PRT; 3015 AA.  
AC Q9PWU9  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.

OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Hepatitis C virus; an infectious molecular clone of a second major  
RT genotype (2a) and lack of viability of intertypic 1a and 2a  
RT chimeras";  
RL Virology 262:250-263 (1999).  
DR EMBL; AF177039; AAF01181.1; -;  
DR EMBL; AF177037; AAF01179.1; -;  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSP; P27958; IHEI.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002522; HCV\_capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR002129; Pyridoxal\_dec.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00487; DEXDC; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;

Query Match 98.2%; Score 3553; DB 2; Length 3015;  
Best Local Similarity 97.5%; Pred. No. 2.3e-240;  
Matches 669; Conservative 9; Mismatches 8; Indels 0; Gaps 0;  
  
Qy 1 MAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
Db :|||||  
1030 LAPITAYAAQQTGRLGCGIITSLTGRDNQVEGEVQIVSTATQTFLATCINGVCWTVYHGA 1089  
Qy 61 GTRTIASPKGFVIMQYTNVDQDLVGVWPAQGSRLTCTCGSDLYLVTRHADVIPVRR 120

Db	1090	GTRTIA	SPKGPVTQMTNVDQDLVGNPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR	1149	GO: GO:0005524; F-ATP binding; IEA.
Qy	121	GDSRGL	SPRPISYIKSGSGGPGLLCPAGHAVGIFRAAACTRGVAKAVDFIPVENLETTM	180	GO: GO:0008026; F-ATP-dependent helicase activity; IEA.
Db	1150	GDSRGL	SPRPISYIKSGSGGPGLLCPAGHAVGLFRAAACTRGVAKAVDFIPVENLGTTM	1209	GO: GO:0003723; F-RNA binding; IEA.
Qy	181	RSPVFT	DNSSPPVPOSFQVAHLHPTGSGSKTKVPAVAAQGYKVLVNLNPSVAATLGFG	240	GO: GO:0003236; F-serine-type peptidase activity; IEA.
Db	1210	RSPVFT	DNSSPPVPOSFQVAHLHPTGSGSKTKVPAVAAQGYKVLVNLNPSVAATLGFG	1269	GO: GO:0003198; F-structural molecule activity; IEA.
Qy	241	AYMSKA	HDIPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS	300	GO: GO:0008508; P-proteolysis and peptidolysis; IEA.
Db	1270	AYMSKA	HDIPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS	1329	GO: GO:0006350; P-transcription; IEA.
Qy	301	ILGIGT	VLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEPYGAIPLEVI	360	GO: GO:0019079; P-viral genome replication; IEA.
Db	1330	ILGIGT	VLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEPYGAIPLEVI	1389	GO: GO:0019087; P-viral transformation; IEA.
Qy	361	KGGRHL	IFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVVATDALMTGYT	420	InterPro: IPR00345; CytC_heme_BS.
Db	1390	KGGRHL	IFCHSKKCCDELAALVALGINAVAYRGLDVSVIPPIGDVVVVVATDALMTGYT	1449	InterPro: IPR001410; DEAD/DEAH_N.
Qy	421	GDPSV	IDNCNCTVQTVDPSLDPTFTIETITLPQDAVSRQRTGRGKPGIYRFVAPG	480	InterPro: IPR001545; DEAD/DEAH_N.
Db	1450	GDPSV	IDNCNCTVQTVDPSLDPTFTIETITLPQDAVSRQRTGRGKPGIYRFVAPG	1509	InterPro: IPR002522; HCV capsid.
Qy	481	ERPSGM	FSDSVLCECDAGCAWYELTPAETTVRLAYMNTPGIPVQDHLFEWEGVFTGL	540	InterPro: IPR002521; HCV core.
Db	1510	ERPSGM	FSDSVLCECDAGCAWYELTPAETTVRLAYMNTPGIPVQDHLFEWEGVFTGL	1569	InterPro: IPR002519; HCV env.
Qy	541	THIDAH	FLSTKSGENLPVLVAYQATVCARAAPPSPQMWKCLIRLPTLHGPTPLL	600	InterPro: IPR002531; HCV NS1.
Db	1570	THIDAH	FLSTKSGENFPVLVAYQATVCARAAPPSPQMWKCLIRLPTLHGPTPLL	1629	InterPro: IPR000745; HCV NS4a.
Qy	601	YRLGAV	QNEITLTHPTVKYIMTQMSADLEVYVTTWVLVGVLAALAAAYCLSTGCWIVGR	660	InterPro: IPR001490; HCV NS4b.
Db	1630	YRLGAV	QNEITLTHPTVKYIMTQMSADLEVYVTTWVLVGVLAALAAAYCLSTGCWIVGR	1689	InterPro: IPR002868; HCV NS5a.
Qy	661	VVLSGR	PAIIPDREVLYREFDEMEEC 686		InterPro: IPR001650; Helicase_C.
Db	1690	VVLSGR	PAIIPDREVLYREFDEMEEC 1715		InterPro: IPR004109; Peptidase_S29.
RESULT 8					
Q9PNX5		PRELIMINARY;	PRT; 3015 AA.		
AC	Q9PNX5				
DT	01-MAY-2000	(TrEMBLrel. 13, Created)			
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)			
DE		Polyprotein.			
OS		Hepatitis C virus.			
OC		Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC		Hepacivirus.			
OX		NCBI_TaxID=11103;			
RN		[1]			
RP		SEQUENCE FROM N.A.			
RX		MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;			
RA		Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;			
RT		"Hepatitis C virus: an infectious molecular clone of a second major			
RT		genotype (2a) and lack of viability of intertypic 1a and 2a			
RT		chimeras."			
RL		Virology 262:250-263(1999).			
DR		EMBL; AF177040; AAF01182.1; -.			
DR		EMBL; AF177038; AAF01180.1; -.			
DR		PIR; PS0326; PS0326.			
DR		PIR; PS0327; PS0327.			
DR		PIR; PS0328; PS0328.			
DR		HSP; P27958; IHEI.			
DR		GO: GO:0016021; C:integral to membrane; IEA.			
DR		GO: GO:0019028; C:viral capsid; IEA.			
DR		GO: GO:0019031; C:viral envelope; IEA.			

Query Match	98.2%	Score 3553;	DB 2;	Length 3015;
Best Local Similarity	97.5%	Pred. No. 2.3e-240;	Mismatches 8;	Indels 0;
Matches 669;	Conservative	9;	Mismatches 8;	Indels 0;
Qy	1	MAPITAYAQOTRGLLCIITSLTGRDKNQVEGEVQIVSTAAQTFLATCINGVCWTVVHGA	60	
Db	1030	LAPITAYAQOTRGLLCIITSLTGRDKNQVEGEVQIVSTATQTFLATCINGVCWTVVHGA	1089	
Qy	61	GTRTIA	SPKGPVTQMTNVDQDLVGNPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR	120
Db	1090	GTRTIA	SPKGPVTQMTNVDQDLVGNPAPQGSRLTPTCGSSDLYLVTRHADVIPVRRR	1149
Qy	121	GDSRGL	SPRPISYIKSGSGGPGLLCPAGHAVGIFRAAACTRGVAKAVDFIPVENLETTM	180
Db	1150	GDSRGL	SPRPISYIKSGSGGPGLLCPAGHAVGLFRAAACTRGVAKAVDFIPVENLGTTM	1209
Qy	181	RSPVFT	DNSSPPVPOSFQVAHLHPTGSGSKTKVPAVAAQGYKVLVNLNPSVAATLGFG	240
Db	1210	RSPVFT	DNSSPPVPOSFQVAHLHPTGSGSKTKVPAVAAQGYKVLVNLNPSVAATLGFG	1269
Qy	241	AYMSKA	HDIPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS	300
Db	1270	AYMSKA	HDIPNIRTVGRTITTTGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDTATS	1329
Qy	301	ILGIGT	VLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEPYGAIPLEVI	360
Db	1330	ILGIGT	VLDOAETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEPYGAIPLEVI	1389



DB	1686	IVLSGKPAIIPDREVLYQSFDEMEEC	1711	
RESULT 10				
O36609		PRELIMINARY;	PRT;	3011 AA.
ID	O36609;			
AC	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein.			
OS	Hepatitis C virus strain H77.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.			
OX	NCBI_TaxID=63746;			
RN	[1]_TaxID=63746;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H77;			
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;			
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;			
RT	Transcripts from a single full-length cDNA clone of hepatitis C virus			
RT	are infectious when directly transfected into the liver of a			
RT	chimpanzee.;			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).			
DR	EMBL; AF011752; AAB67037.1; -.			
DR	PIR; A44150; A44150.			
DR	PIR; PQ0804; PQ0804.			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P26664; IHEI.			
DR	GO; GO:0019028; C:Viral capsid; IEA.			
DR	GO; GO:0019031; C:Viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006350; P:transcription; IEA.			
DR	GO; GO:0019079; P:viral genome replication; IEA.			
DR	GO; GO:0019087; P:viral transformation; IEA.			
DR	InterPro; IPR00345; CytC_heme_BS.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002522; HCV_capsid.			
DR	InterPro; IPR002521; HCV_core.			
DR	InterPro; IPR002519; HCV_env.			
DR	InterPro; IPR002531; HCV_NS1.			
DR	InterPro; IPR000745; HCV_NS4a.			
DR	InterPro; IPR001490; HCV_NS4b.			
DR	InterPro; IPR002868; HCV_NS5a.			
DR	InterPro; IPR002166; HCV_RdRP.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR009003; Pept_Ser_Cys.			
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.			
DR	InterPro; IPR007095; RNA_pol_DS_PS.			
DR	InterPro; IPR007094; RNA_pol_PSVir.			
DR	Pfam; PF01543; HCV_capsid; 1.			
DR	Pfam; PF01542; HCV_core; 1.			
DR	Pfam; PF01539; HCV_env; 1.			
DR	Pfam; PF01560; HCV_NS1; 1.			
DR	Pfam; PF01538; HCV_NS2; 1.			
DR	Pfam; PF02907; HCV_NS3; 1.			
DR	Pfam; PF01006; HCV_NS4a; 1.			
DR	Pfam; PF01001; HCV_NS4b; 1.			
DR	Pfam; PF01506; HCV_NS5a; 1.			
DR	Pfam; PF00271; Helicase_C; 1.			
DR	Pfam; PF00998; Viral_RdRP; 1.			
DR	SMART; SM00487; DEXDC; 1.			
DR	PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.			
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;				
KW Polyprotein; Transmembrane.				
SQ SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA9B5DC CRC64;				
Query Match 97.9%; Score 3542; DB 2; Length 3011;				
Best Local Similarity 97.2%; Pred. No. 1.4e-239;				
Matches 667; Conservative 9; Mismatches 10; Indels 0; Gaps 0;				
QY	1	MAPITAYAAQOTRGLGCIITSLTGRDKQVEGEVQIVSTAAOTFLATCINGCVTVYHGA	60	
DB	1026	LAPITAYAAQOTRGLGCIITSLTGRDKQVEGEVQIVSTATQTLATCINGCVTVYHGA	1085	
QY	61	GTRTTASPGKPVIOYNTVDLVGMPAPQSGRSITPCTCGSSDLYLVTRHADVIPRRR	120	
DB	1086	GTRTTASPGKPVIOYNTVDLVGMPAPQSGRSITPCTCGSSDLYLVTRHADVIPRRR	1145	
QY	121	GDSRGLSPRPISYLVKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM	180	
DB	1146	GDSRGLSPRPISYLVKSGSGGPGLLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM	1205	
QY	181	RSPVFTDSSPPVPPQSFQVAHLHAPTGGKSTKVPAAVAAQGYKVLVLPNSVAATLFG	240	
DB	1206	RSPVFTDSSPPVPPQSFQVAHLHAPTGGKSTKVPAAVAAQGYKVLVLPNSVAATLFG	1265	
QY	241	AYMSKAHGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATS	300	
DB	1266	AYMSKAHGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATS	1325	
QY	301	ILIGITVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI	360	
DB	1326	ILIGITVLDOAETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYGKAIPLEVI	1385	
QY	361	KGRHLIFCHSKKKCDLAALVALGINAVAYRGLDVSVIPIGDVVVVVATDALMTGYT	420	
DB	1386	KGRHLIFCHSKKKCDLAALVALGINAVAYRGLDVSVIPIGDVVVVVATDALMTGYT	1445	
QY	421	GDPSVIDNCTQTVDFSLDPTFTIETILPQDAVSRTRGRTRGKPGIYRFVAPG	480	
DB	1446	GDPSVIDNCTQTVDFSLDPTFTIETILPQDAVSRTRGRTRGKPGIYRFVAPG	1505	
QY	481	ERPSGMFSDSVLCECYDAGCAMEYLTPTAETTVRLRAYMNTPGLPVQCDHLEFMEGVFTGL	540	
DB	1506	ERPSGMFSDSVLCECYDAGCAMEYLTPTAETTVRLRAYMNTPGLPVQCDHLEFMEGVFTGL	1565	
QY	541	THIDAHFLSQTQSGENLPIYVAYQATVCARAQAPPPSDQMKCLIRLKLHGPTPL	600	
DB	1566	THIDAHFLSQTQSGENLPIYVAYQATVCARAQAPPPSDQMKCLIRLKLHGPTPL	1625	
QY	601	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWLVGGVLAALAAAYCLSTGCWIVGR	660	
DB	1626	YRLGAVQNEITLTHPVTKYIMTCMSADLEVTSTWLVGGVLAALAAAYCLSTGCWIVGR	1685	
QY	661	VVLSGKPAIIPDREVLYREFDEMEEC	686	
DB	1686	IVLSGKPAIIPDREVLYREFDEMEEC	1711	
RESULT 11				
O36610		PRELIMINARY;	PRT;	3011 AA.
ID	O36610			
AC	036610;			
DT	01-JAN-1998 (TrEMBLrel. 05, Created)			
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein.			
OS	Hepatitis C virus strain H77.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.			
OX	NCBI_TaxID=63746;			
RN	[1]_TaxID=63746;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H77;			
RC	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;			
RX				

RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
 RT "Transcripts from a single full-length cDNA clone of hepatitis C virus  
 are infectious when directly transfected into the liver of a  
 chimpanzee".  
 RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).  
 RL EMBL: AF011753; AAB67038.1; -;  
 DR PIR: A4150; A4150.  
 DR PIR: PQ0804; PQ0804.  
 DR PIR: PS0326; PS0326.  
 DR PIR: PS0327; PS0327.  
 DR PIR: PS0328; PS0328.  
 DR HSP; P27958; lHEI.  
 DR GO: GO:0019028; C: viral capsid; IEA.  
 DR GO: GO:0019031; C: viral envelope; IEA.  
 DR GO: GO:0005524; F: ATP binding; IEA.  
 DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.  
 DR GO: GO:0003723; F: RNA binding; IEA.  
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F: serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F: structural molecule activity; IEA.  
 DR GO: GO:0006350; P: proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P: transcription; IEA.  
 DR GO: GO:0019079; P: viral genome replication; IEA.  
 DR GO: GO:0019087; P: viral transformation; IEA.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR011545; DEAD/DEAH\_N.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRp.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR004109; Peptidase\_S29.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR InterPro: IPR002518; Pept\_U39\_HCV NS2.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral RdRp; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR Coar protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; Transmembrane.  
 SQ SEQUENCE 3011 AA; 327222 MW; 293P91430A0D4067 CRC64;

Query Match 97.9%; Score 3542; DB 2; Length 3011;  
 Best Local Similarity 97.4%; Pred. No. 1.4e-239;  
 Matches 667; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 2 APITAYAOQTGGLGCIITSLTGDKQVEGEVQIVSTAAOTFLATCINGVCWTVYHGAG 61  
 DB 1027 APITAYOQTGGLGCIITSLTGDKQVEGEVQIVSTATOTFLATCINGVCWTVYHGAG 1086  
 QY 62 TRTIASPKGPVIQMYTNVDQLVGNPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRRG 121  
 DB 1087 TRTIASPKGPVIQMYTNVDQLVGNPAPQGSRSITPCTCGSSDLYLVTRHADVIPVRRRG 1146  
 QY 122 DSRGSLSPRISYIKGSGGFLICPAGHAGVIFRAAVCTRGKAVKADFIPIVENLTTMR 181  
 DB 1147 DSRGSLSPRISYIKGSGGFLICPAGHAGVIFRAAVCTRGKAVKADFIPIVENLTTMR 1206

QY 182 SPVFTDNSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAQAQYKVLVLPNSVAATLGFGA 241  
 DB 1207 SPVFTDNSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAQAQYKVLVLPNSVAATLGFGA 1266  
 QY 242 YMSKAGHIDPNIRITGVRTITGSPITYSTYVKFLADGGCSGGAYDIIICDECHSTDATSI 301  
 DB 1267 YMSKAGHIDPNIRITGVRTITGSPITYSTYVKFLADGGCSGGAYDIIICDECHSTDATSI 1326  
 QY 302 LGIGTVLDQAETAGARLVVLATATPPGSSVTVPHNIEEVALSTTGTGEPFYKGAIPLEVIK 361  
 DB 1327 LGIGTVLDQAETAGARLVVLATATPPGSSVTVPHNIEEVALSTTGTGEPFYKGAIPLEVIK 1386  
 QY 362 GGRHLIFCHSKKCDLAELAKLVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYTG 421  
 DB 1387 GGRHLIFCHSKKCDLAELAKLVALGINAVAYRGLDVSVPPIGDVVVATDALMTGYTG 1446  
 QY 422 DFDSDVDCNTCVTQTVDPSLDPTFTTITLTPQDAVSRRTORRGTRGKPGIYRFVAPGE 481  
 DB 1447 DFDSDVDCNTCVTQTVDPSLDPTFTTITLTPQDAVSRRTORRGTRGKPGIYRFVAPGE 1506  
 QY 482 RPSGMFDSVLCEDYDAGCAWYELTPTAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLT 541  
 DB 1507 RPSGMFDSVLCEDYDAGCAWYELTPTAETTVRLRAYMNTPLPVCQDHLFEWGVFTGLT 1566  
 QY 542 HIDAHLFSQTKSGENLPYLVAQVATVCARAQAPPPSWDQWKLIHLKPTLHGPTLLY 601  
 DB 1567 HIDAHLFSQTKSGENLPYLVAQVATVCARAQAPPPSWDQWKLIHLKPTLHGPTLLY 1626  
 QY 602 RLGAQVNEITLTHPVTKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGRV 661  
 DB 1627 RLGAQVNEITLTHPVTKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGRV 1686  
 QY 662 VLSGKPAIIPDREVLVREFDEMEEC 686  
 DB 1687 VLSGKPAIIPDREVLVREFDEMEEC 1711

RESULT 12  
 Q03463 ID Q03463 PRELIMINARY; PRT; 3011 AA.  
 AC Q03463;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Polyprotein.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92044440; PubMed=1658196;  
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,  
 Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated  
 from a human carrier: comparison with reported isolates for conserved  
 and divergent regions";  
 RT J. Gen. Virol. 72:2697-2704 (1991).  
 RL (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93117120; PubMed=1335573;  
 RA Okamoto H., Kanai N., Mishihiro S.;  
 RT "Full-length nucleotide sequence of a Japanese hepatitis C virus  
 isolate (HC-J1) with high homology to USA isolates";  
 RL Nucleic Acids Res. 20:6410-6410 (1992).  
 RL (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91013116; PubMed=2170712;  
 RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,  
 Yoshizawa H., Tada F., Miyakawa Y., Mayumi M.;  
 RT "The 5'-terminal sequence of the hepatitis C virus genome";  
 RL Jpn. J. Exp. Med. 60:167-177 (1990).

[4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94174722; PubMed=7510436;  
 RA Mink M., Benichou S., Madale P., Tiollais P., Prince A.,  
 RA Inchauspe G.;  
 RT "Characterization and mapping of a B-cell immunogenic domain in  
 RT hepatitis C virus E2 glycoprotein using a yeast peptide library.";  
 RL Virology 200:246-255(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Okamoto H.;  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: D10749; BAA01582.1; -;  
 DR PIR: PS0326; PS0326.  
 DR PIR: PS0327; PS0327.  
 DR PIR: PS0328; PS0328.  
 DR PIR: S40770; S40770.  
 DR HSSP: P26664; 1HEI.  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0019031; C:viral envelope; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006350; P:transcription; IEA.  
 DR GO: GO:0019079; P:viral genome replication; IEA.  
 DR GO: GO:0019087; P:viral transformation; IEA.  
 DR InterPro: IPR000345; CytC\_heme\_BS.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR011545; DEAD/DEAH\_N.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002166; HCV NS5a.  
 DR InterPro: IPR002166; HCV NS5a.  
 DR InterPro: IPR001650; Helicase\_C29.  
 DR InterPro: IPR004109; Peptidase\_S29.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR InterPro: IPR002518; Pept\_U3\_HCV\_NS2.  
 DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro: IPR007094; RNA\_pol\_PSVir.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase\_C; 1.  
 DR Pfam: PF00998; Viral\_RDRP; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CVTOCHROME\_C; UNKNOWN 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; Transmembrane.  
 SQ SEQUENCE 3011 AA; 347114 MW; 97E9052C0250463B CRC64;  
 Query Match 97.9%; Score 3542; DB 2; Length 3011;  
 Best Local Similarity 96.9%; Pred. No. 1.4e-239;  
 Matches 665; Conservative 10; Mismatches 11; Indels 0; Gaps 0;  
 Qy 1 MAPITAVAQTRGLGCIITSLTRDRKNQVEGEVQIVSTAAQTFLATCINGCVCTVYHGA 60  
 Db :|||||  
 1026 LAPITAVAQTRGLGCIITSLTRDRKNQVEGEVQIVSTAAQTFLATCINGCVCTVYHGA 1085

QY 61 GTRTIASPKGPVIQMYTNVDQDLVGNPAPQGGSRSLTPTCGSSDLYLVTRHADVIPVRR 120  
 Db |||||  
 1086 GTRTIASPKGPVIQMYTNVDQDLVGNPAPQGGSRSLTPTCGSSDLYLVTRHADVIPVRR 1145  
 QY 121 GDSRGLSPRDISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVENLETTM 180  
 Db |||||  
 1146 GDSRGLSPRDISYLVKSSGGPLCPAGHAGVIFRAAVCTRGVAKAVDFIPVESLETTM 1205  
 QY 181 RSPVFTDNSSPPVPQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPNPSVAATLGF 240  
 Db |||||  
 1206 RSPVFTDNSSPPAVPQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPNPSVAATLGF 1265  
 QY 241 AYMSKAHGIDPNIRGTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDTATS 300  
 Db |||||  
 1266 AYMSKAHGIDPNIRGTITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDTATS 1325  
 QY 301 ILGIGTVLDOAETAGARLVVLTATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEVI 360  
 Db |||||  
 1326 VLIGIGTVLDOAETAGARLVVLTATPPGSIIVPHNIEEVALSTTGEIPFYGKAIPLEAI 1385  
 QY 361 KGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSVIPPIGDVVVVVATDALMTGYT 420  
 Db |||||  
 1386 KGRHLIFCHSKKKCDLAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYT 1445  
 QY 421 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLPQDAVSRTQRRGTGRGKPGIYRFVAPG 480  
 Db |||||  
 1446 GDFDSVIDCNTCTQTVDFSLDPTFTIETITLPQDAVSRTQRRGTGRGKPGIYRFVAPG 1505  
 QY 481 ERPSGMFDSVLCCEYDAGCANYELTPAETTVRLRAYMNTPGIPVCCODHLEFWEVFTGL 540  
 Db |||||  
 1506 ERPSGMFDSVLCCEYDAGCANYELTPAETTVRLRAYMNTPGIPVCCODHLEFWEVFTGL 1565  
 QY 541 THIDAHFLSQTQSGENLPVYAYQATVCARAAQPPSWDQMKCLIRLKPHTLHGPTPLL 600  
 Db |||||  
 1566 THIDAHFLSQTQSGENFPYVAYQATVCARAAQPPSWDQMKCLIRLKPHTLHGPTPLL 1625  
 QY 601 YRLGAVQNEITLTHPVTKYIMTSCMSADLEVTSTWLVGVLAALAAAYCLSTGCVVIVGR 660  
 Db |||||  
 1626 YRLGAVQNEITLTHPVTKYIMTSCMSADLEVTSTWLVGVLAALAAAYCLSTGCVVIVGR 1685  
 QY 661 VVLSGRPAIIPDBREVLYREFDEMEEC 686  
 Db |||||  
 1686 VVLSGRPAIIPDBREVLYREFDEMEEC 1711  
 RESULT 13  
 Q9DIT6  
 ID Q9DIT6 PRELIMINARY; PRT; 3011 AA.  
 AC Q9DIT6  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Genomic RNA for polyprotein gene.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21014672; PubMed=11115058;  
 RA Kumar U., Tuthill T., Thomas H.C., Monjardino J.;  
 RA "Sequence, expression and reconstitution of an HCV genome from a  
 RT British isolate derived from a single blood donation.";  
 RL J. Viral Hepat. 7:459-465(2000).  
 DR EMBL: AJ278830; CAC03609.1; -;  
 DR PIR: PS0326; PS0326.  
 DR PIR: PS0327; PS0327.  
 DR PIR: PS0328; PS0328.  
 DR HSSP: P27958; 1A1V.  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0019028; C:viral capsid; IEA.  
 DR GO: GO:0019031; C:viral envelope; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.



DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
 DR GO: 0003723; F: RNA binding; IEA.  
 DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
 DR GO: 0008236; F: serine-type peptidase activity; IEA.  
 DR GO: 0005198; F: structural molecule activity; IEA.  
 DR GO: 0006508; P: proteolysis and peptidolysis; IEA.  
 DR GO: 0006350; P: transcription; IEA.  
 DR GO: 0019079; P: viral genome replication; IEA.  
 DR GO: 0019087; P: viral transformation; IEA.  
 DR GO: 000345; CytC heme BS.  
 DR InterPro: IPR001410; DEAD/DEAH\_N.  
 DR InterPro: IPR011545; DEAD/DEAH\_N.  
 DR InterPro: IPR011492; Flavi\_DEAD.  
 DR InterPro: IPR002522; HCV capsid.  
 DR InterPro: IPR002521; HCV core.  
 DR InterPro: IPR002519; HCV env.  
 DR InterPro: IPR002531; HCV NS1.  
 DR InterPro: IPR000745; HCV NS4a.  
 DR InterPro: IPR001490; HCV NS4b.  
 DR InterPro: IPR002868; HCV NS5a.  
 DR InterPro: IPR002166; HCV RdRP.  
 DR InterPro: IPR001650; Helicase C.  
 DR InterPro: IPR004109; Peptidase S29.  
 DR InterPro: IPR009003; Pept\_Ser\_Cys.  
 DR InterPro: IPR002518; Pept\_U39 HCV NS2.  
 DR InterPro: IPR007095; RNA pol DS PS.  
 DR InterPro: IPR007094; RNA pol PSvir.  
 DR Pfam: PF07652; Flavi\_DEAD; 1.  
 DR Pfam: PF01543; HCV capsid; 1.  
 DR Pfam: PF01542; HCV core; 1.  
 DR Pfam: PF01539; HCV env; 1.  
 DR Pfam: PF01560; HCV NS1; 1.  
 DR Pfam: PF01538; HCV NS2; 1.  
 DR Pfam: PF02907; HCV NS3; 1.  
 DR Pfam: PF01006; HCV NS4a; 1.  
 DR Pfam: PF01001; HCV NS4b; 1.  
 DR Pfam: PF01506; HCV NS5a; 1.  
 DR Pfam: PF00271; Helicase C; 1.  
 DR Pfam: PF00998; Viral RdRP; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 DR Polyprotein; Transmembrane.  
 KW CHAIN 810 1026 non-structural protein 2.  
 FT CHAIN 1027 1657 non-structural protein 3.  
 FT CHAIN 1658 1711 non-structural protein 4a.  
 FT CHAIN 1712 191 core protein.  
 FT CHAIN 1912 1972 non-structural protein 4b.  
 FT CHAIN 1973 2420 non-structural protein 5a.  
 FT CHAIN 2421 3011 non-structural protein 5b.  
 FT CHAIN 192 383 envelop protein 1.  
 FT CHAIN 384 809 envelop protein 2.  
 SQ SEQUENCE 3011 AA; 327406 MW; 786264A74A5452D3 CRC64;  
 Query Match 97.8%; Score 3540; DB 2; Length 3011;  
 Best Local Similarity 97.1%; Pred. No. 1.9e-239;  
 Matches 666; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MAPITAAQQTGRLGCGIITSLTGRDKNQVEGEQIVSTAQTFLATCINGCVTVYHGA 60  
 DB LAPIITAAQQTGRLGCGIITSLTGRDKNQVEGEQIVSTAQTFLATCINGCVTVYHGA 1085  
 QY 61 GTRTASPKGVQIYNTVDDQVGPAPQGRSLTPTCTGSSDLYLVRHADVIPVRRR 120  
 DB GTRTASPKGVQIYNTVDDQVGPAPQGRSLTPTCTGSSDLYLVRHADVIPVRRR 1145  
 QY 121 GDSRGLSPRPISYLKSGSGPLCPAGHVGIFRAAVCTRGVAKAVDFIPVENLETTM 180  
 DB GDSRGLSPRPISYLKSGSGPLCPAGHVGIFRAAVCTRGVAKAVDFIPVENLETTM 1205  
 QY 181 RSPVFTDNSSPPVPSQVQVHLHAPTSGSKTKVPAAYAAQGYKVLVNFVAATLGF 240  
 DB RSPVFTDNSSPPVPSQVQVHLHAPTSGSKTKVPAAYAAQGYKVLVNFVAATLGF 1265

QY 241 AYMSKAHGIDPNIRTTGSPITTYSTYKFLADGGCGGAYDIIICDECHSDATS 300  
 DB AYMSKAHGIDPNIRTTGSPITTYSTYKFLADGGCGGAYDIIICDECHSDATS 1325  
 QY 301 ILGIGTVLDOAETAGARLVVLTATPPGSTVTPHPNIEEVALSTTGIPFYGKAIPLEVI 360  
 DB ILGIGTVLDOAETAGARLVVLTATPPGSTVTPHPNIEEVALSTTGIPFYGKAIPLEVI 1385  
 QY 361 KGRHLIFCHSKKKCDLAAKVALGINAVAYRGLDVSVIPGIDVVVATDALMTGYT 420  
 DB KGRHLIFCHSKKKCDLAAKVALGINAVAYRGLDVSVIPGIDVVVATDALMTGYT 1445  
 QY 421 GDFDSVIDCNTCTQTVDPSLDTFTIETTLTQDAVSRTQRRGRTGRGKPIRVFVAPG 480  
 DB GDFDSVIDCNTCTQTVDPSLDTFTIETTLTQDAVSRTQRRGRTGRGKPIRVFVAPG 1505  
 QY 481 ERPSGMPDSSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLBFMEGVFTGL 540  
 DB ERPSGMPDSSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCDHLBFMEGVFTGL 1565  
 QY 541 THIDAHFSLQKSGENLPYLVAQVATVCARAQPPSWDQMKCLIRLKPTLHGPTPL 600  
 DB THIDAHFSLQKSGENLPYLVAQVATVCARAQPPSWDQMKCLIRLKPTLHGPTPL 1625  
 QY 601 YRLGAVONEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCLSTGCVWIVGR 660  
 DB YRLGAVONEITLTHPTVKYIMTMSADLEVTSTWLVGGVLAALAAAYCLSTGCVWIVGR 1685  
 QY 661 VVLSGKPAIIPDREVLRYREFDEMEEC 686  
 DB IVLSGKPAIIPDREVLRYREFDEMEEC 1711  
 RESULT 14  
 POLG\_HCVH STANDARD; PRT; 3011 AA.  
 ID POLG\_HCVH  
 AC P27958;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)  
 DE NSB (P27); Nonstructural protein NS4A (P4); Nonstructural protein  
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TaxID=11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92052256; PubMed=1658800;  
 RA Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,  
 RA Prince A.M.;  
 RT "Genomic structure of the human prototype strain H of hepatitis C  
 RT virus: comparison with American and Japanese isolates.";  
 RL proc. Natl. Acad. Sci. U.S.A. 88:10292-10296 (1991).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.";  
 RL Nat. Struct. Biol. 4:463-467 (1997).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493270; DOI=10.1016/S0969-2126(98)00010-0;  
 RA Kim J.L., Morgerstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,  
 RA Murcko M.A., Lin C., Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound  
 RT oligonucleotide: the crystal structure provides insights into the mode

of unwinding.";  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: Protease NS2 is responsible for the cleavage of NS2-NS3.  
 CC -!- FUNCTION: Protease NS3 is responsible for the cleavage of NS3-NS4A. NS4A-NS4B, NS4B-NS5A and NS5A-NS5B.  
 CC -!- FUNCTION: NS4A forms a complex with NS3 and is essential for the activation of NS3.  
 CC -!- FUNCTION: NS5A seems to have a transcriptional activatory role.  
 CC -!- FUNCTION: NS5B is a RNA-dependent RNA polymerase that plays an essential role in the virus replication.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).  
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: E1 and E2. The nucleocapsid is a complex of protein C and mRNA.  
 CC -!- PTM: The structural proteins C, E1 and E2 are produced by proteolytic processing by the host signal peptidases.  
 CC -!- SIMILARITY: Contains 1 peptidase S29 domain.  
 CC -!- SIMILARITY: Contains 1 peptidase U39 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M67463; AAA45534.1; -;  
 CC PIR; A36814; GNMVCH.  
 CC DR PDB; 1A1R; X-ray; A/B=1017-1214, C/D=1676-1698.  
 CC DR PDB; 1A1V; X-ray; A=1192-1667.  
 CC DR PDB; 1HEI; X-ray; -;  
 CC DR MEROPS; S29.001; -;  
 CC DR MEROPS; U39.001; -;  
 CC DR TRANSFAC; T04155; -;  
 CC DR InterPro; IPR001410; DEAD.  
 CC DR InterPro; IPR002522; HCV capsid.  
 CC DR InterPro; IPR002521; HCV Core.  
 CC DR InterPro; IPR002519; HCV env.  
 CC DR InterPro; IPR002531; HCV NS1.  
 CC DR InterPro; IPR000745; HCV NS4a.  
 CC DR InterPro; IPR001490; HCV NS4b.  
 CC DR InterPro; IPR002868; HCV NS5a.  
 CC DR InterPro; IPR002166; HCV RdRp.  
 CC DR InterPro; IPR001650; Helicase\_C.  
 CC DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 CC DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 CC DR InterPro; IPR004109; Peptidase\_S29.  
 CC DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 CC DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 CC DR Pfam; PF01543; HCV\_capsid; 1.  
 CC DR Pfam; PF01542; HCV\_core; 1.  
 CC DR Pfam; PF01539; HCV\_env; 1.  
 CC DR Pfam; PF01560; HCV\_NS1; 1.  
 CC DR Pfam; PF01538; HCV\_NS2; 1.  
 CC DR Pfam; PF02907; HCV\_NS3; 1.  
 CC DR Pfam; PF01006; HCV\_NS4a; 1.  
 CC DR Pfam; PF01001; HCV\_NS4b; 1.  
 CC DR Pfam; PF01506; HCV\_NS5a; 1.  
 CC DR Pfam; PF00271; Helicase\_C; 1.  
 CC DR Pfam; PF00998; Viral RdRp; 1.  
 CC DR SMART; SM00487; DEXDC; 1.  
 CC 3D-structure: ATP-binding; Coat protein; Core protein;  
 KW Envelope protein; Glycoprotein; Helicase; Hydrolase;  
 KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 KW Serine protease; Transferase; Transmembrane.  
 FT INIT\_MET 1 1 Removed from capsid protein C by the  
 FT cellular aminopeptidase.  
 FT CHAIN 1 191 Capsid protein C.

FT CHAIN 192 383 Envelope glycoprotein E1.  
 FT CHAIN 384 746 Envelope glycoprotein E2.  
 FT CHAIN 747 809 Protein P7.  
 FT CHAIN 810 1026 Nonstructural protein NS2.  
 FT CHAIN 1027 1657 Protease/helicase NS3.  
 FT CHAIN 1658 1711 Nonstructural protein NS4A.  
 FT CHAIN 1712 1972 Nonstructural protein NS4B.  
 FT CHAIN 1973 2420 Nonstructural protein NS5A.  
 FT CHAIN 2421 3011 Nonstructural protein NS5B.  
 FT CHAIN 3012 369 Potential.  
 FT TRANSMEM 370 369 Charge relay system (By similarity).  
 FT ACT\_SITE 1083 1083 Charge relay system (By similarity).  
 FT ACT\_SITE 1107 1107 Charge relay system (By similarity).  
 FT ACT\_SITE 1165 1165 Charge relay system (By similarity).  
 FT NP\_BIND 1230 1237 ATP (Potential).  
 FT SITE 1316 1319 DECH box.  
 FT CARBOHYD 196 196 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 209 209 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 234 234 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 305 305 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 417 417 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 423 423 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 430 430 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 448 448 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 476 476 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 532 532 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 540 540 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 556 556 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 576 576 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 623 623 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 645 645 N-linked (GlcNAc. .) (Potential).  
 FT STRAND 1032 1035  
 FT HELIX 1039 1048  
 FT STRAND 1050 1050  
 FT STRAND 1059 1063  
 FT STRAND 1068 1074  
 FT TURN 1075 1076  
 FT STRAND 1077 1081  
 FT HELIX 1082 1085  
 FT TURN 1086 1087  
 FT STRAND 1090 1092  
 FT TURN 1093 1094  
 FT STRAND 1095 1097  
 FT STRAND 1101 1103  
 FT TURN 1104 1107  
 FT STRAND 1108 1112  
 FT TURN 1115 1116  
 FT STRAND 1120 1120  
 FT STRAND 1122 1122  
 FT STRAND 1129 1133  
 FT TURN 1135 1136  
 FT STRAND 1139 1144  
 FT TURN 1147 1148  
 FT STRAND 1149 1157  
 FT HELIX 1158 1161  
 FT TURN 1162 1163  
 FT TURN 1165 1166  
 FT STRAND 1168 1170  
 FT TURN 1172 1173  
 FT STRAND 1176 1186  
 FT TURN 1187 1188  
 FT STRAND 1189 1197  
 FT HELIX 1198 1205  
 FT STRAND 1680 1687  
 FT STRAND 1693 1694  
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 97.1%; Score 3513; DB 1; Length 3011;  
 Best Local Similarity 96.6%; Pred. No. 1-5e-237;  
 Matches 663; Conservative 10; Mismatches 13; Indels 0; Gaps 0;  
 Qy 1 MAPITAYAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 60  
 Db 1026 LAPITAYAQTRGLGCIITSLTGRDNQKQVEGEVQIVSTAAQTFLATCINGVCWTVYHGA 1085



Db	1266	AYMSKAHVDPNIRTVRTITTCAPITYTYGKFLADGGCGGAYDIIICDECHSTDSTS	1325
Qy	301	ILGIGTVLQDAETAGARLVVLTATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIEPLEVI	360
Db	1326	ILGIGTVLQDAETAGARLVVLTATATPPGSVTVPHPNIEEVALSNTGEIPFYGKAIEPIETI	1385
Qy	361	KGRHLIFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVPPIGDVVVVVATDALMTGYT	420
Db	1386	KGRHLIFCHSKKKCDLAAKLVALGINAVAYYRGLDVSVPPIGDVVVVVATDALMTGYT	1445
Qy	421	GPDFSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOREGRTRGKPGIYRFVAPG	480
Db	1446	GPDFSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTOREGRTRGKPGIYRFVTPG	1505
Qy	481	ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCQDHLFEWEGVFTGL	540
Db	1506	ERPSGMFDSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVQCQDHLFEWESVFTGL	1565
Qy	541	THIDAHFLSOTKOSGENLPYLVAIQATVCARAQAPPPSWDMWKCLIRLKPTLHGPTPLL	600
Db	1566	THIDAHFLSOTKQAGNFPYLVAIQATVCARAQAPPPSWDMWKCLIRLKPTLHGPTPLL	1625
Qy	601	YRLGAVQNEITLTHPVTKYIMTCSADLEVVTTWVLVGGVLAALAAAYCLSTGCVVIVGR	660
Db	1626	YRLGAVQNDVTLTHPVTKFIMACMSADLEVVTTWVLVGGVLAALAAAYCLTSGSVIVGR	1685
Qy	661	VVLSGKPAIIPDREVLYREFDEMEEC	686
Db	1686	IILSGRPAIIPDREVLYQEFDEMEEC	1711

Search completed: November 7, 2005, 20:16:35  
Job time : 91.9151 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 19:57:16 ; Search time 95.7758 Seconds  
(without alignments)  
3347.654 Million cell updates/sec

Title: US-10-658-782-4

Perfect score: 4455

Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRSTGKSWGKPGYPWP 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4455	100.0	829	5	Aae18690 Multiple
2	4455	100.0	829	7	Adc06769 Chimeric
3	4455	100.0	829	8	Adl66807 HCV multi
4	4032	90.5	1099	5	Aau76378 HCV multi
5	4032	90.5	1099	6	Abg72262 HCV multi
6	4032	90.5	1099	8	Adl66809 HCV multi
7	2222	49.9	1021	2	Aaw34481 HCV anti
8	2222	49.9	1021	2	Aaw40039 Fusion pr
9	2222	49.9	1021	5	Aae22050 pSOD/c200
10	1627.5	36.5	1766	1	Aap92041 Hepatitis
11	1624.5	36.5	2261	1	Aap90164 Peptide e
12	1624.5	36.5	2436	1	Aap92050 HCV prote
13	1624.5	36.5	2436	1	Aap90288 Peptide e
14	1624.5	36.5	2772	3	Aab18540 Protein e
15	1624.5	36.5	2772	8	Adn35976 HCV cdna
16	1624.5	36.5	2955	2	Aay14975 Amino aci
17	1624.5	36.5	2955	3	Aab18541 Polypept
18	1624.5	36.5	2955	8	Adn35978 HCV cdna
19	1624.5	36.5	3011	2	Aar21519 Completed
20	1624.5	36.5	3011	2	AAR90931 Hepatitis
21	1624.5	36.5	3011	2	Aaw40038 HCV polyp
22	1624.5	36.5	3011	2	Aaw40038 HCV polyp
23	1624.5	36.5	3011	5	Aae22049 Hepatitis
24	1624.5	36.5	3011	8	Adl23107 Hepatitis
25	1624.5	36.5	3011	8	Adr29357 Hepatitis

#### ALIGNMENTS

##### RESULT 1

ID	AAE18690	standard; protein; 829 AA.
AC	AAE18690;	
XX		
DT	17-MAY-2002	(first entry)
DE	Multiple epitope fusion antigen (MEFA) 12 protein.	
XX		
KW	Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;	
KW	HCV infection; MEFA 12 protein.	
XX		
OS	Unidentified.	
XX		
EH	Key	Location/Qualifiers
FT	Misc-difference 315	/note= "Encoded by ATG"
FT	Misc-difference 645	/note= "Encoded by GAG"
XX		
PN	WO200196875-A2.	
XX		
PD	20-DEC-2001.	
XX		
XX	14-JUN-2001; 2001WO-US019369.	
XX		
PR	15-JUN-2000; 2000US-0212082P.	
PR	02-APR-2001; 2001US-0280811P.	
PR	02-APR-2001; 2001US-0280867P.	
XX		
XX	(CHIR ) CHIRON CORP.	
XX		
PI	Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;	
PI	Medina-Selby A;	
XX		
DR	WPI; 2002-179522/23.	
DR	N-PSDB; AAD29796.	
XX		
PT	Immunoassay solid support useful for detecting hepatitis C virus	
PT	infection in a biological sample, comprises at least one of HCV anti-core	
PT	antibody and HCV NS3/4a epitope, bound to the support.	
XX		
PS	Disclosure; Fig 7; 87pp; English.	
XX		
CC	The present invention relates to hepatitis C virus (HCV) core antigen and	
CC	NS (nonstructural) 3/4a antibody combination assay that can detect both	
CC	HCV antigens and antibodies present in a sample using a single solid	

Aar22208 Sequence  
Aar21565 HCV CKS-3  
Aar33632 HCV CKS-3  
Aar33594 HCV CKS-3  
Aar33574 HCV CKS-3  
Aar52690 HCV CKS-3  
Aab51172 HCV recom  
Abw01858 HCV-CKS-3  
Aau84597 HCV polyp  
Aap92047 HCV prote  
Aar08123 Hepatitis  
Aar24440 Composite  
Aar34009 HCV-1 pol  
Aar25135 HCV polyp  
Aar70230 Composite  
Aar31621 Hepatitis  
Aar28582 HCV amino  
Aap90158 Protein a  
Aar08124 Hepatitis  
Aar38451 Hepatitis

26 1623 36.4 781 2 AAR22208  
27 1623 36.4 781 2 AAR21565  
28 1623 36.4 781 2 AAR33632  
29 1623 36.4 781 2 AAR33594  
30 1623 36.4 781 2 AAR33574  
31 1623 36.4 781 2 AAR52690  
32 1623 36.4 781 2 AAB51172  
33 1623 36.4 781 4 AABW01858  
34 1622.5 36.4 3011 5 AAU84597  
35 1621.5 36.4 2301 1 AAP92047  
36 1621.5 36.4 2772 2 AAR08123  
37 1619.5 36.4 2894 2 AAR24440  
38 1616.5 36.3 2816 2 AAR34009  
39 1615.5 36.3 2435 2 AAR25135  
40 1614.5 36.2 2894 2 AAR70230  
41 1614.5 36.2 3011 2 AAR31621  
42 1608.5 36.1 2436 2 AAR28582  
43 1605.5 36.0 1786 1 AAP90158  
44 1600.5 35.9 2955 2 AAR08124  
45 1597.5 35.9 1986 8 ADR38451

CC matrix as well as immunoassay solid supports for use in the assay. The  
CC solid support is useful for detecting HCV infection in a biological  
CC sample. The present sequence is MEFA (multiple epitope fusion antigen) 12  
CC protein. This sequence is used in the exemplification of the invention  
XX  
SQ Sequence 829 AA;  
Query Match 100.0%; Score 4455; DB 5; Length 829;  
Best Local Similarity 100.0%; Pred. No. 2.4e-310; Indels 0; Gaps 0;  
Matches 829; Conservative 0; Mismatches 0;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Db 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Qy 121 AARTTSLTSLFSPGASONIQLITSTDNSSPPVQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Db 121 AARTTSLTSLFSPGASONIQLITSTDNSSPPVQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Qy 181 AQGYKVLVLPNSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGC 240  
Db 181 AQGYKVLVLPNSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGC 240  
Qy 241 SGGAYDIIICDECHSTATSIIGTGLDQETAGARLVLTATTPGCVTVPHNIEEV 300  
Db 241 SGGAYDIIICDECHSTATSIIGTGLDQETAGARLVLTATTPGCVTVPHNIEEV 300  
Qy 301 ALSTTGIEIPFYGKAIPLVETKGGHLLFCHSKKCDLAAKLVALGINAVAYRGLDVS 360  
Db 301 ALSTTGIEIPFYGKAIPLVETKGGHLLFCHSKKCDLAAKLVALGINAVAYRGLDVS 360  
Qy 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCASCGRPAIIPDREVLRYEFDEMECSQH 420  
Db 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCASCGRPAIIPDREVLRYEFDEMECSQH 420  
Qy 421 LPYIEQGMMLAEQFKQALGLSRGKPAIVPDKVLVQYDDEMECSQAAPYIEQAQVIA 480  
Db 421 LPYIEQGMMLAEQFKQALGLSRGKPAIVPDKVLVQYDDEMECSQAAPYIEQAQVIA 480  
Qy 481 HQPEKVLGLIDNDQVVVTPDKETLYEAFDDEMECSKAALIEGQRMALMSKIOGLL 540  
Db 481 HQPEKVLGLIDNDQVVVTPDKETLYEAFDDEMECSKAALIEGQRMALMSKIOGLL 540  
Qy 541 GILRRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDYNPLV 600  
Db 541 GILRRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDYNPLV 600  
Qy 601 ETWKKPDYEPVVGRRSRRAQALPVWARPDYNPLVETWKKPDYEPVVGRRKTKRNT 660  
Db 601 ETWKKPDYEPVVGRRSRRAQALPVWARPDYNPLVETWKKPDYEPVVGRRKTKRNT 660  
Qy 661 NRRPDQVKFGGGQIVGGVYLLPRRGLVLATRTKTSPIPKARRPGRGRTWAQPGYWP 720  
Db 661 NRRPDQVKFGGGQIVGGVYLLPRRGLVLATRTKTSPIPKARRPGRGRTWAQPGYWP 720  
Qy 721 YGNKDRSTGKSGKPGYWPMPKRTKRNTRPDQVFPFGGQIVGGVYLLPRRGLVL 780  
Db 721 YGNKDRSTGKSGKPGYWPMPKRTKRNTRPDQVFPFGGQIVGGVYLLPRRGLVL 780  
Qy 781 ATRKTSIPKARRPGRGRTWAQPGYWPVLYGNKDRSTGKSGKPGYWP 829  
Db 781 ATRKTSIPKARRPGRGRTWAQPGYWPVLYGNKDRSTGKSGKPGYWP 829  
RESULT 2  
ADC06769 ID ADC06769 standard; protein; 829 AA.  
XX  
AC ADC06769;

XX 18-DEC-2003 (first entry)  
DT Chimeric multiple epitope fusion antigen 12 protein.  
DE  
XX immunoassay solid support; HCV; NS3/4a; non-structural;  
KW non-A, non-B hepatitis; NANB; multiple epitope fusion antigen 12; MEFA12;  
KW chimeric.  
XX Chimeric.  
OS Synthetic.  
OS Unidentified.  
OS Hepatitis C virus.  
OS Homo sapiens.  
XX US2002192639-A1.  
PN 19-DEC-2002.  
XX 14-JUN-2001; 2001US-00881239.  
PP 15-JUN-2000; 2000US-0212082P.  
XX 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280857P.  
PR (CHIE/) CHIEN D Y.  
XX (ARCA/) ARCANGEL P.  
PA (TAND/) TANDESKE L.  
PA (GEOR/) GEORGE-NASCIMENTO C.  
PA (COIT/) COIT D.  
PA (MEDI/) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX WPI; 2003-644609/61.  
DR N-PSDB; ADC06770.  
XX  
PT Immunoassay solid support for detecting hepatitis C virus infection in  
PT biological samples, comprises a hepatitis C virus anti-core antibody and  
PT an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
PT antibody.  
XX Claim 45; Fig 7; 40pp; English.  
XX The invention relates to a novel immunoassay solid support comprising at  
CC least one hepatitis C virus (HCV) anti-core antibody and at least one  
CC isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereco.  
CC The system of the invention may be useful for detecting HCV infection in  
CC a biological sample and for treating or detecting non-A, non-B hepatitis  
CC (NANB hepatitis). The current sequence is that of the chimeric multiple  
CC epitope fusion antigen 12 (MEFA12) protein of the invention.  
XX  
SQ Sequence 829 AA;  
Query Match 100.0%; Score 4455; DB 7; Length 829;  
Best Local Similarity 100.0%; Pred. No. 2.4e-310;  
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Db 61 AGPHFNPLSTRGCNCSYPGHITGHRMAWKLGSAAARTTSGFVSLFAPGAKQNETHTVTGGA 120  
Qy 121 AARTTSLTSLFSPGASONIQLITSTDNSSPPVQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Db 121 AARTTSLTSLFSPGASONIQLITSTDNSSPPVQSFQVAHLHAPTGSKGSTKVPAAAYA 180  
Qy 181 AQGYKVLVLPNSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGC 240  
Db 181 AQGYKVLVLPNSVAATLGFCAYSKAHGIDPNIRTVRTITTTGSPITYSTYKFLADGGC 240

Db 181 AOGKVLVLPNSVAATLGFAGYMSKAHGDIPNIRTVRTTTTGGSPITTYTGKFLADGGC 240  
Qy 241 SGGAYDIIICDECHSTDATSIIGTGLDQAEAGARLVVLATATPPGSVTVPHNIEEV 300  
Db 241 SGGAYDIIICDECHSTDATSIIGTGLDQAEAGARLVVLATATPPGSVTVPHNIEEV 300  
Qy 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV 360  
Db 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV 360  
Qy 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
Db 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
Qy 421 LPYTEQGMMLAEQFKQKALGLSRGKPAIIPDKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
Db 421 LPYTEQGMMLAEQFKQKALGLSRGKPAIIPDKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
Qy 481 HQFKEKVLGLINDQVVVTPDKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
Db 481 HQFKEKVLGLINDQVVVTPDKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
Qy 541 GILRRHVGPGEVAVQMMNRLIAFASRGNHVSPTHVPSRRRFAQALPVWARPDPNPLV 600  
Db 541 GILRRHVGPGEVAVQMMNRLIAFASRGNHVSPTHVPSRRRFAQALPVWARPDPNPLV 600  
Qy 601 ETWKKPDYEPVPHGRSSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHGRKTKRNT 660  
Db 601 ETWKKPDYEPVPHGRSSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHGRKTKRNT 660  
Qy 661 NRRPDQVKFGGQIIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRWAOFGYPMPL 720  
Db 661 NRRPDQVKFGGQIIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRWAOFGYPMPL 720  
Qy 721 YGNKDRSTGSKGKPGYPWPRKTKRNTNRRPDQVKFGGQIIVGGVYLLPRGPRGLV 780  
Db 721 YGNKDRSTGSKGKPGYPWPRKTKRNTNRRPDQVKFGGQIIVGGVYLLPRGPRGLV 780  
Qy 781 ATRKTSPIPKARRPEGRWAOFGYPMPLYGKDRRSTGSKGKPGYPMW 829  
Db 781 ATRKTSPIPKARRPEGRWAOFGYPMPLYGKDRRSTGSKGKPGYPMW 829  
RESULT 3  
ADL66807  
ID ADL66807 standard; protein; 829 AA.  
XX AC ADL66807;  
XX DT 03-JUN-2004 (first entry)  
XX DE HCV multiple epitope fusion antigen 12 (MEFA 12) polypeptide.  
XX KW HCV; MEFA 12; HCV antigen; HCV polyprotein;  
KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection;  
KW multiple epitope fusion antigen 12.  
XX OS Hepatitis C virus.  
XX PN WO2004021871-A2.  
XX PD 18-MAR-2004.  
XX PF 08-SEP-2003; 2003WO-US028071.  
XX PR 09-SEP-2002; 2002US-0409515P.  
XX PA (CHIR ) CHIRON CORP.  
XX PI Arcangel P, Chien D;  
XX WPI; 2004-248333/23.  
DR N-PSDB; ADL66806.

XX Detecting hepatitis C virus (HCV) infection in a biological sample by  
PT detecting complexes formed between the HCV antibody and the antigens from  
PT the first region of the HCV polyprotein and the multiple epitope fusion  
PT antigen (MEFA).  
XX Claim 14; SEQ ID NO 4; 93pp; English.  
XX The invention relates to a method of detecting hepatitis C virus (HCV)  
CC infection in a biological sample. The method comprises providing an  
CC immunassay solid support comprising HCV antigens bound to it, where the  
CC HCV antigens comprise one or more isolated antigens form a first region  
CC of the HCV polyprotein, combining a biological sample with the solid  
CC support under conditions that allow HCV antibodies, when present in the  
CC biological sample, to bind to the one or more HCV antigens, adding to the  
CC solid support a detectably labelled HCV multiple epitope fusion antigen  
CC (MEFA), where the labelled MEFA comprises at least one epitope from the  
CC same region of the HCV polyprotein as the one or more isolated antigens,  
CC where the MEFA binds to the bound HCV antibody, and detecting complexes  
CC formed between the HCV antibody and the one or more antigens from the  
CC first region of the HCV polyprotein and the MEFA, if any, as an  
CC indication of HCV infection in the biological sample. The method is  
CC useful for detecting hepatitis C virus (HCV) infection in a biological  
CC sample. This sequence represents the MEFA 12 polypeptide used in the  
CC scope of the invention.  
XX  
SQ Sequence 829 AA;  
Query Match 100.0%; Score 4455; DB 8; Length 829;  
Best Local Similarity 100.0%; Pred. No. 2.4e-310;  
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MATKAVCLVKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
Db 1 MATKAVCLVKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLGHFVHFGDNTAGCTS 60  
Qy 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPCAKONETHVTGGA 120  
Db 61 AGPHFNPLSTRGCNCSYIPGHIITGHRMAWKLGSAAARTTSGFVSLFAPCAKONETHVTGGA 120  
Qy 121 AARTTSGTSLFSPGASQNIQLITSTONSPPVPPQSFQVAHLHAPTSGSKSTKVPAAAYA 180  
Db 121 AARTTSGTSLFSPGASQNIQLITSTONSPPVPPQSFQVAHLHAPTSGSKSTKVPAAAYA 180  
Qy 181 AOGYKVLVLPNSVAATLGFAGYMSKAHGDIPNIRTVRTTTTGGSPITTYTGKFLADGGC 240  
Db 181 AOGYKVLVLPNSVAATLGFAGYMSKAHGDIPNIRTVRTTTTGGSPITTYTGKFLADGGC 240  
Qy 241 SGGAYDIIICDECHSTDATSIIGTGLDQAEAGARLVVLATATPPGSVTVPHNIEEV 300  
Db 241 SGGAYDIIICDECHSTDATSIIGTGLDQAEAGARLVVLATATPPGSVTVPHNIEEV 300  
Qy 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV 360  
Db 301 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV 360  
Qy 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
Db 361 IPTSGDVVVATDALMTGYTGDPSVIDCNTCAGSKPAIIPDREVLVYREFDEMECSQH 420  
Qy 421 LPYTEQGMMLAEQFKQKALGLSRGKPAIIPDKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
Db 421 LPYTEQGMMLAEQFKQKALGLSRGKPAIIPDKEVLYQQYDEMECSQAAPYIEQAQVIA 480  
Qy 481 HQFKEKVLGLINDQVVVTPDKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
Db 481 HQFKEKVLGLINDQVVVTPDKEILYAFDEMECSKAALIEGQRMALMLKSKIQGLL 540  
Qy 541 GILRRHVGPGEVAVQMMNRLIAFASRGNHVSPTHVPSRRRFAQALPVWARPDPNPLV 600  
Db 541 GILRRHVGPGEVAVQMMNRLIAFASRGNHVSPTHVPSRRRFAQALPVWARPDPNPLV 600  
Qy 601 ETWKKPDYEPVPHGRSSRRFAQALPVWARPDPNPLVETWKKPDYEPVPHGRKTKRNT 660

Db 601 ETWKKPDYEPVHVGRSSRFAQALPVWARPDPYNPPLVETWKKPDYEPVHVGRKTKRNT 660  
Qy 661 NRRPQDVKFGGQIVGGVLLPRRGRLGLVATRKTSPIPKARRPGRGTWQPGYWPPL 720  
Db 661 NRRPQDVKFGGQIVGGVLLPRRGRLGLVATRKTSPIPKARRPGRGTWQPGYWPPL 720  
Qy 721 YGNKDRRSTGKSGKPGYWPMPKRTKRNTRRPQDVKFGGQIVGGVLLPRRGRLGLV 780  
Db 721 YGNKDRRSTGKSGKPGYWPMPKRTKRNTRRPQDVKFGGQIVGGVLLPRRGRLGLV 780  
Qy 781 ATRKTSPIPKARRPGRGTWQPGYWPPLYGNKDRRSTGKSGKPGYWPMP 829  
Db 781 ATRKTSPIPKARRPGRGTWQPGYWPPLYGNKDRRSTGKSGKPGYWPMP 829

RESULT 4  
AAU76378  
ID AAU76378 standard; protein; 1099 AA.  
XX  
AC AAU76378;  
DT 08-MAY-2002 (first entry)  
XX  
DE HCV multiple epitope fusion antigen (MEFA) 7.1 protein sequence.  
XX  
KV Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;  
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;  
KW non-structural protein.  
XX  
OS Hepatitis C virus.  
OS Synthetic.  
XX  
PN W0200196870-A2.  
XX  
PD 20-DEC-2001.  
XX  
XX 14-JUN-2001; 2001WO-US019156.  
XX  
PR 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
XX  
XX WPI; 2002-090228/12.  
DR N-PSDB; ABK15345.  
XX  
XX  
PT Immunoassay solid support, useful for detecting hepatitis C virus  
PT infection in biological sample, comprises HCV NS3/4a conformational  
PT epitope and multiple epitope fusion antigen bound to the support.  
XX  
XX Claim 5; Fig 5; 92pp; English.  
XX  
XX The present invention relates to a new immunoassay solid support  
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a  
CC conformational epitope and a multiple epitope fusion antigen (MEFA),  
CC bound to the support. The NS3/4a conformational epitope and/or MEFA  
CC reacts specifically with anti-HCV antibodies present in a biological  
CC sample from an HCV-infected individual. The immunoassay of the invention  
CC is useful for detecting hepatitis C virus infection in a biological  
CC sample. The method of the invention provides a sensitive, accurate  
CC diagnostic and prognostic tool to provide adequate patient care and to  
CC prevent transmission of HCV by blood and by blood products, or by  
CC personal contact. Use of NS3/4a conformational epitope in combination  
CC with MEFA, provides a sensitive and reliable method for detecting early  
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing  
CC masking problems, improving sensitivity in detecting antibodies by  
CC allowing a greater number of epitopes on a unit surface area of  
CC substrate, and improving substrate. Detection accuracy is increased and

CC the incidence of false results is reduced because of the identification  
CC and the use of highly immunogenic HCV antigens which are present during  
CC the early stages of HCV seroconversion. The present amino acid sequence  
CC represents the multiple epitope fusion antigen (MEFA) 7.1 of the  
CC invention  
XX  
SQ Sequence 1099 AA;  
Query Match 90.5%; Score 4032; DB 5; Length 1099;  
Best Local Similarity 69.8%; Pred. No. 7.6e-280;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFEGDNTAGCTS 60  
Qy 61 AGPHFNPPLSTR----- 71  
Db 61 AGPHFNPPLSRKHGKPKDEERHVGDLGNVTADKGVADSVIEDSVISLGDHCIIGRTLTV 120  
Qy 72 -----GCNCSIYPGHITGHRMAWKLS 93  
Db 121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIAQNLSGNCNCSIYPGHITGHRMAWKLS 180  
Qy 94 AARTTSGFVSLFAPGAKQNETHTVTGAAARTTSLTSLFSPGASQNIQLITS----- 145  
Db 181 AARTTSGFVSLFAPGAKQNETHTVTGAAARTTSLTSLFSPGASQNIQLIVDFIPVENLE 240  
Qy 146 -----TDNSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPSPAATL 197  
Db 241 TTMRSPVFTDNSSPPVVPQSFQVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPSPAATL 300  
Qy 198 GFGAYMSKAHGIDPNIRITGVTITGSPITYSTYKFLADGGCGGAYDIIICDECHSTD 257  
Db 301 GFGAYMSKAHGIDPNIRITGVTITGSPITYSTYKFLADGGCGGAYDIIICDECHSTD 360  
Qy 258 ATSLGIGTVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTGTBIPFYGKAIP 317  
Db 361 ATSLGIGTVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTGTBIPFYGKAIP 420  
Qy 318 EVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSITPSGDVVVATDALMT 377  
Db 421 EVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSITPSGDVVVATDALMT 480  
Qy 378 GYTGDPSVIDCNTC----- 392  
Db 481 GYTGDPSVIDCNTCTQTQVDFSLDPTFTTITLPODAVSRQRRGRTGRGKPGIYRFV 540  
Qy 393 ----- 392  
Db 541 APCRPSGMPDSSVLCBCEYDAGCAWVELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVF 600  
Qy 393 ----- 392  
Db 601 TGLTHIDAHFLSQTOSGENLPYLVAQATVCARAQAPPPSDMMKCLIRLKLPTLHGPT 660  
Qy 393 -----ACSGKPAIIPDREVLYREFDSEEE 416  
Db 661 PLLIYRLGAVQNEITLTHPTVKYIMTCSADLEVVVTSACSKKPAIIPDREVLYREFDSEEE 720  
Qy 417 CSQHLPIYEQGMMLAQFKOKALGLSRGKPKAIPVQKEVLYQQYDEMEECSSQAAPYIEQA 476  
Db 721 CSQHLPIYEQGMMLAQFKOKALGLSRGKPKAIPVQKEVLYQQYDEMEECSSQAAPYIEQA 780  
Qy 477 QVIAHQFKEKVLGLINDQVWVTPDKIILYEAFDEMEECASKALITEGORMAEMLSKI 536  
Db 781 QVIAHQFKEKVLGLINDQVWVTPDKIILYEAFDEMEECASKALITEGORMAEMLSKI 840  
Qy 537 QGLLGILRRHVGPGEAVQMMNRLIAFASGNHVSPTHYVPSRSRRFAQALPVWARPDPYN 596  
Db 841 QGLLGILRRHVGPGEAVQMMNRLIAFASGNHVSPTHYVPSRSRRFAQALPVWARPDPYN 900  
Qy 597 PPLVETWKKPDYEPVHVGRSSRFAQALPVWARPDPYNPPLVETWKKPDYEPVHVGRKTK 656



Db 901 PPLVETWKKEDYBPVGHSSRRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRKT 960  
Qy 657 KRNTNRPDQKPPGGQIVGGVLLPRRGLVLAIRKTSPTPKARRPEGRWAOQGY 716  
Db 961 KRNTNRPDQKPPGGQIVG-----RRGP-----FIPKARRPEGRWAOQGY 1003  
Qy 717 PWPLYGNKDRRSTGSKGKGYPMWRKTKRNTNRPDQKPPGGQIVGGVLLPRRGR 776  
Db 1004 PWPLYGNKDRRSTGSKGKGYPMWRKTKRNTNRPDQKPPGGQIVG-----RRGP- 1056  
Qy 777 LGVLATRKTSPIPKARRPEGRWAOQGYPMWRKTKRNTNRPDQKPPGGQIVG 829  
Db 1057 -----PIPKARRPEGRWAOQGYPMWRKTKRNTNRPDQKPPGGQIVG 1099

RESULT 5  
ABG72262  
ID ABG72262 standard; protein; 1099 AA.  
XX ABG72262;  
XX AC  
XX ABG72262;  
DT 06-MAR-2003 (first entry)  
XX HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).  
DE  
XX  
XX  
XX Immunassay solid support; Hepatitis C Virus type-1; HCV-1; HCV-2;  
KW NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;  
KW MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;  
KW HCV infection; Hepatitis C Virus type-2; Hepatitis C Virus type-3;  
KW mutant; mutain.  
XX  
OS Hepatitis C virus type 1.  
OS Hepatitis C virus type 2.  
OS Hepatitis C virus type 3.  
OS Synthetic.  
OS Chimeric.

Key Location/Qualifiers  
FH 1. .156  
FT Region  
FT /note= "Correspond to amino acids 1-156 of HCV-1 HSD  
FT superoxide dismutase")  
FT Region  
FT /note= "Correspond to amino acids 303-320 of HCV-1 E1"  
FT Region  
FT /note= "Correspond to consensus sequence of amino acids  
FT 179. .199  
FT 390-410 of HCV-1 E2 HVR"  
FT Region  
FT /note= "Correspond to consensus sequence of amino acids  
FT 384-414 of HCV-1 and HCV-2 E2 HVR"  
FT Region  
FT /note= "Correspond to amino acids 1193-1658 of HCV-1  
FT helicase"  
FT 699. .745  
FT /note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
FT 1 epitope"  
FT Region  
FT /note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
FT 748. .794  
FT /note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
FT 1 epitope"  
FT Region  
FT /note= "Correspond to amino acids 1689-1735 of HCV-2 5-1-  
FT 797. .843  
FT /note= "Correspond to amino acids 1689-1735 of HCV-2 5-1-  
FT 1 epitope"  
FT Region  
FT /note= "Correspond to amino acids 1901-1936 of HCV-1  
FT polypeptide C100"  
FT 846. .881  
FT /note= "Correspond to amino acids 1901-1936 of HCV-1  
FT polypeptide C100"  
FT Region  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT 884. .919  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT Region  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT 922. .957  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT Region  
FT /note= "Correspond to core region antigenic determinants  
FT 958. .1028  
FT /note= "Correspond to core region antigenic determinants

FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
FT 1029. .1099  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
XX  
PN US2002146685-A1.  
XX 10-OCT-2002.  
XX 14-JUN-2001; 2001US-00881654.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIE/) CHIEN D Y.  
PA (ARCA/) ARCANGEL P.  
PA (TAND/) TANDESKE L.  
PA (GEOR/) GEORGE-NASCIMENTO C.  
PA (COIT/) COIT D.  
PA (MEDI/) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
PI Medina-Selby A;  
PI WPI; 2003-147573/14.  
DR N-PSDB; ABX14411.  
DR  
XX Immunassay solid support for detecting Hepatitis C Virus infection in  
FT biological samples, comprises Hepatitis C Virus conformational epitope  
FT and multiple epitope fusion antigen.  
XX Claim 25; Fig 5A-5F; 45pp; English.  
XX The present invention relates to immunoassays comprising Hepatitis C  
CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
CC multiple epitope fusion antigen react with anti-HCV antibodies present in  
CC a biological sample from an HCV-infected individual. The immunoassays and  
CC methods of the invention are useful for detecting HCV infection in a  
CC biological sample. The inventive immunoassay solid support provides a  
CC sensitive and reliable method for detecting early HCV seroconversion. The  
CC assays can detect HCV infection caused by any six known genotypes of HCV.  
CC The use of the multiple epitope fusion proteins decreases masking  
CC problems, improves sensitivity in detecting antibodies by allowing a  
CC greater number of epitopes on a unit area of substrate, and improves  
CC selectivity. The present sequence represents HCV multiple epitope fusion  
CC antigen 7.1 (MEFA 7.1), a mutant HCV polypeptide derived from various  
CC regions of HCV type 1, 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide  
CC sequences  
XX  
SQ Sequence 1099 AA;  
Query Match 90.5%; Score 4032; DB 6; Length 1099;  
Best Local Similarity 69.8%; Pred. No. 7.6e-280;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHNPLSTR----- 71  
Db 61 AGPHNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCIIIGRLTW 120  
Qy 72 -----GNCSTIYPGHITGHRMAWKLS 93  
Db 121 HEKADDLKGGNEESTXTGNAGSLACGVIGIAQLNLSGNCSTIYPGHITGHRMAWKLS 180  
Qy 94 AARTTSFVSLFAPCAKQNETHVITGGAARTTSLTSLFSPGASQNTLITS----- 145

Db	181	AARTSGFVSLFAPGAKQNETHTVGAAARTTSLTSLFSPGASQNIQLIVDFIPVENLE	240
Qy	146	-----TDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL	197
Db	241	TTWRSPPVFTDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL	300
Qy	198	GFGAYMSKAHGIDPNIRLTGRTITTTGSPITYSYTGKFLADGGCGGAYDIICDECHSTD	257
Db	301	GFGAYMSKAHGIDPNIRLTGRTITTTGSPITYSYTGKFLADGGCGGAYDIICDECHSTD	360
Qy	258	ATSILGIGTVLQDAETAGARLVVLTATPPGTVTPHPNIEEVALSTTGEIPIPGYKAIFL	317
Db	361	ATSILGIGTVLQDAETAGARLVVLTATPPGTVTPHPNIEEVALSTTGEIPIPGYKAIFL	420
Qy	318	EVIKGRHLIFCHSKKKCCDELAALKVALGINAVAYYRGLDVSVIPSGDVVVVATDALMT	377
Db	421	EVIKGRHLIFCHSKKKCCDELAALKVALGINAVAYYRGLDVSVIPSGDVVVVATDALMT	480
Qy	378	GYTGFDSVIDCNTC-----	392
Db	481	GYTGFDSVIDCNTCTQTQVDFSLDPTFTIETITLPQDAVSRTQRRGRTGRGPGIYRFV	540
Qy	393	-----	392
Db	541	APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFPWEGVF	600
Qy	393	-----	392
Db	601	TGLTHIDAHPLSQTQSGENLPYLVAQATVCARAQAPPSPQDMWKILRLKTLHGPT	660
Qy	393	-----ACSGKPAIIPDREVLXREFDEME	416
Db	661	PLYLRLGAVQNETLTHPVTKYIMTMSADLEVTACSCKPALIIPDREVLXREFDEME	720
Qy	417	CSOHLPIYIEGMMLABQFKQKALGLSRGKPAIVDPKEVLYQYDEMECSQAAPYIEQA	476
Db	721	CSOHLPIYIEGMMLABQFKQKALGLSRGKPAIVDPKEVLYQYDEMECSQAAPYIEQA	780
Qy	477	QVIAHOFKEKVLGLINDQVVVTPDKELIYEAFDEMEECASKAALIEEGORMAEMLSKI	536
Db	781	QVIAHOFKEKVLGLINDQVVVTPDKELIYEAFDEMEECASKAALIEEGORMAEMLSKI	840
Qy	537	QGLGLILRRHVGEGAVQWMNRLIAPASRGNHVSPTHYVPSRRRPAQALPVMWAPDYN	596
Db	841	QGLGLILRRHVGEGAVQWMNRLIAPASRGNHVSPTHYVPSRRRPAQALPVMWAPDYN	900
Qy	597	PPLVETWKKPDYBPPVVHGRSSRRFAQALPVMWAPDYNPPLVETWKKPDYBPPVHGRKT	656
Db	901	PPLVETWKKPDYBPPVVHGRSSRRFAQALPVMWAPDYNPPLVETWKKPDYBPPVHGRKT	960
Qy	657	KRNTNRRPDQVKPFGGQIVGGVYLLPRGPRGLVLAIRKTSPIPKARRPEGRTWAPGY	716
Db	961	KRNTNRRPDQVKPFGGQIVG-----RRGP-----PIPKARRPEGRTWAPGY	1003
Qy	717	PWPLYGNKDRSRCKSGKPGYWPBKRTKNTNRRPDQVKPFGGQIVGGVYLLPRGPR	776
Db	1004	PWPLYGNKDRSRCKSGKPGYWPBKRTKNTNRRPDQVKPFGGQIVG-----RRGP-1056	
Qy	777	LGVLATRKTSPIPKARRPEGRTWAPGYWPPLYGNKDRSRCKSGKPGYWP 829	
Db	1057	-----PIPKARRPEGRTWAPGYWPPLYGNKDRSRCKSGKPGYWP 1099	
RESULT 6			
ADL66809			
ID	ADL66809	standard; protein; 1099 AA.	
XX	AC	ADL66809;	
XX	DT	03-JUN-2004 (first entry)	
XX	DE	HCV multiple epitope fusion antigen 7.1 (MEFA 7.1) polypeptide.	

KW	HCV; MEFA 7.1; HCV antigen; HCV polyprotein;
KW	multiple epitope fusion antigen; MEFA; hepatitis C virus infection;
KW	multiple epitope fusion antigen 7.1.
OS	Hepatitis C virus.
XX	
PN	WO2004021871-A2.
XX	
PD	18-MAR-2004.
XX	
PF	08-SEP-2003; 2003WO-US028071.
XX	
PR	09-SEP-2002; 2002US-0409515P.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
PI	Arcangel P, Chien D;
XX	
DR	WPI; 2004-248333/23.
DR	N-PSDB; ADL66808.
XX	
PT	Detecting hepatitis C virus (HCV) infection in a biological sample by
PT	detecting complexes formed between the HCV antibody and the antigens from
PT	the first region of the HCV polyprotein and the multiple epitope fusion
PT	antigen (MEFA).
XX	
PS	Claim 15; SEQ ID NO 6; 93pp; English.
XX	
CC	The invention relates to a method of detecting hepatitis C virus (HCV)
CC	infection in a biological sample. The method comprises providing an
CC	immunoassay solid support comprising HCV antigens bound to it, where the
CC	HCV antigens comprise one or more isolated antigens form a first region
CC	of the HCV polyprotein, combining a biological sample with the solid
CC	support under conditions that allow HCV antibodies, when present in the
CC	biological sample, to bind to the one or more HCV antigens, adding to the
CC	solid support a detectably labelled HCV multiple epitope fusion antigen
CC	(MEFA), where the labelled MEFA comprises at least one epitope from the
CC	same region of the HCV polyprotein as the one or more isolated antigens,
CC	where the MEFA binds to the bound HCV antibody, and detecting complexes
CC	formed between the HCV antibody and the one or more antigens from the
CC	first region of the HCV polyprotein and the MEFA, if any, as an
CC	indication of HCV infection in the biological sample. The method is
CC	useful for detecting hepatitis C virus (HCV) infection in a biological
CC	sample. This sequence represents the MEFA 7.1 polypeptide used in the
CC	scope of the invention.
XX	
SQ	Sequence 1099 AA;
Query Match 90.5%; Score 4032; DB 8; Length 1099;	
Best Local Similarity 69.8%; Pred. No. 7.6e-280;	
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;	
QY	1 MATKAVCVLKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGHVHFGDNTAGCTS 60
DB	1 MATKAVCVLKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGHVHFGDNTAGCTS 60
QY	61 AGHENPLSTR----- 71
DB	61 AGHENPLSRKHGPKDEERHVDLGNVTRADKGVADVSDISLSDHCLIGRTLIV 120
QY	72 -----GCNCSIYPGHITGHRMAWKLS 93
DB	121 HEKADDLKGGNESTKTGNAGSRLACGVIGIAQLNLSGNCNCSIYPGHITGHRMAWKLS 180
QY	94 AARTTSGFVSLFAPGAKQNETHTVGAAARTTSLTSLFSPGASQNIQLITS----- 145
DB	181 AARTTSGFVSLFAPGAKQNETHTVGAAARTTSLTSLFSPGASQNIQLIVDFIPVENLE 240
QY	146 -----TDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL 197
DB	241 TTWRSPPVFTDNSSPPVPSQFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLNPSVAATL 300
QY	198 GFGAYMSKAHGIDPNIRITGRTITTTGSPITYSYTGKFLADGGCGGAYDIICDECHSTD 257



Db	115	GRTLVLVHEKADDLGKGGNEESTK-TGNAGSRLACGVI-----GIAQNLEFGAVDFIPVEN	168
Qy	142	LITS-----TDNSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	195
Db	169	LETTWRSVPFTDSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	228
Qy	196	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYTKFLADGGCGGAYDIIICDECHS	255
Db	229	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYTKFLADGGCGGAYDIIICDECHS	288
Qy	256	TDATSLTGIGTGLDQAEAGARLVLATATPGSVTVPHNPINEEVALSTTGEIPFYGKAI	315
Db	289	TDATSLTGIGTGLDQAEAGARLVLATATPGSVTVPHNPINEEVALSTTGEIPFYGKAI	348
Qy	316	PLEVIKGGRHLLIFCHSKKKCDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	375
Db	349	PLEVIKGGRHLLIFCHSKKKCDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	408
Qy	376	MTGYTGDFDSVIDCNTC-----	392
Db	409	MTGYTGDFDSVIDCNTCVTQTVDLSLDTFTTETITLPQDAVSRTORRGTRGKPGIYR	468
Qy	393	-----	392
Db	469	FVAPGERPSGMFDSVLCBYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLSEFWEG	528
Qy	393	-----	392
Db	529	VFTGLTHIDAHFLSQTQKSGENLPYLVAQATVCARAQAAPPSPSDQMWKCLIRLKPTLHG	588
Qy	393	-----	392
Db	589	PTPLLYRLGAVQNEITLTPVTKYIMTCSADLEVTSTWLVGGVLAALAAVCLSTGCV	648
Qy	393	-----ACSGKPAIIPDREVLRYREFDEMBCSOHLPIYEQGMLAEQFKALGL-----S	442
Db	649	VIVGRVLSGKPAIIPDREVLRYREFDEMBCSOHLPIYEQGMLAEQFKALGLLQTA	708
Qy	443	RGKPAIVPDKEVLYQYD-----EMBECSQAAPYIEQAQVIAHQFKEKVLGLIINDQVV	497
Db	709	RQAE-VIAPAVQTNWQKLETFWAKHMWNFISGIQYLAGLSTLPG--NPATASLMAFTAAV	765
Qy	498	VTP-----DKEILYE-----AFDEMEECASKAALIEBQORMAEMLKSKIQ	537
Db	766	TSPLTTSTQTLLENLILGWVAAQAAPGAATAFVGAGLAGAAGSVGLGKVLIDILAGYGA	825
Qy	538	GLLG-----ILRRHVGFGEVAVQWVN	558
Db	826	GVAGALVAFKIMGEVPSDEDLVNLIPAILSPGALVGVVCAALLRRHVGFGEVAVQWVN	885
Qy	559	RLIAFASRGNHVSPTTHYVPSRSRRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRSS	618
Db	886	RLIAFASRGNHVS-----GNSS	903
Qy	619	RRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRKTRNRRPQDVKPPGCGQIVGG	678
Db	904	T-----NP-----KPKQ-----KKNKNTNRPPQDVKPPGCGQIVGG	934
Qy	724	VYLLPRGPRLGLVLAIRKTS-----PIPKARRPGRWTAQPGYWPPLYGNK-----	724
Db	935	VYLLPRGPRLGLVLAIRKTSERSQPRRRQPIPKARRPGRWTAQPGYWPPLYGNCGCW	994
Qy	725	-----DRRSKSGWKPKPGYWPMPKTRKN	747
Db	995	AGWLLSPRGRSPSGPTD---PRRRSRN	1019
RESULT 8			
AAW40039			
ID AAW40039 standard; protein; 1021 AA.			
XX			
AC AAW40039;			
XX			

DT	26-MAY-1998	(first entry)	
XX	Fusion protein c200/c22.		
DE	Hepatitis C virus C domain; HCV; immunological activity; c200/c22;		
XX	NS3 domain; NS4 domain; S domain; NS5 domain; fusion protein.		
KW	Synthetic.		
OS	Hepatitis virus.		
XX	US5712087-A.		
PN	27-JAN-1998.		
XX	12-MAY-1995; 95US-00440519.		
XX	04-APR-1990; 90US-00504352.		
PR	07-JUL-1992; 92US-00910760.		
XX	(CHIR ) CHIRON CORP.		
PA	Kuo G, Houghton M, Choo Q;		
XX	WPI; 1998-119973/11.		
DR	N-PSDB; AAV09990.		
XX	Immunosays for hepatitis C virus antibodies - using combinations of		
PT	antigenic fragments of HCV polyprotein.		
XX	Example 6; Fig 4; 59pp; English.		
CC	This sequence represents a fusion protein constructed from the hepatitis		
CC	C virus core domain (which is situated at the carboxy terminus of the		
CC	fusion protein) and a c200 construct (a fusion of the NS3 and NS3		
CC	domains). This protein used in the construction of novel combinations of		
CC	HCV antigens that have a broader range of immunological activity than any		
CC	single HCV antigen. An example of such an antigen given in this		
CC	specification comprises a first antigen containing at least 8 amino acids		
CC	of the C domain of the HCV polyprotein and a second antigen comprising at		
CC	least 8 amino acids of the NS3 domain, the NS4 domain, the S domain or		
CC	the NS5 domain of the HCV polyprotein in the form of a fusion protein, a		
CC	physical mixture or bound to a solid matrix		
XX	Sequence 1021 AA;		
Qy	Query Match	49.9%; Score 2222; DB 2; Length 1021;	
Db	Best Local Similarity	46.2%; Pred. No. 4.1e-150;	
XX	Matches 512; Conservative	35; Mismatches 111; Indels 450; Gaps 22;	
Qy	1	MATKAVCVLKGDPVQGIINFQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS	60
Db	1	MATKAVCVLKGDPVQGIINFQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS	60
Qy	61	AGPHFNPLSRGNCISYPGHITGHRNAWKLSAARTTSG-----FVSL-----	104
Db	61	AGPHFNPLSRK-----HGGPKDEERHVGDLGNVTADKDGADVSDSIEDSVISLGDHCII	114
Qy	105	-----FAPGAKQNEHTVTCGAAARTTSGLTSLFSPGASQNIQ-----	141
Db	115	GRTLVLVHEKADDLGKGGNEESTK-TGNAGSRLACGVI-----GIAQNLEFGAVDFIPVEN	168
Qy	142	LITS-----TDNSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	195
Db	169	LETTWRSVPFTDSSPPVVPQSFQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVLPNSVAA	228
Qy	196	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYTKFLADGGCGGAYDIIICDECHS	255
Db	229	TLGFGAYMSKAHGIDPNIRTGVRTITTTGSPITYSTYTKFLADGGCGGAYDIIICDECHS	288
Qy	256	TDATSLTGIGTGLDQAEAGARLVLATATPGSVTVPHNPINEEVALSTTGEIPFYGKAI	315
Db	289	TDATSLTGIGTGLDQAEAGARLVLATATPGSVTVPHNPINEEVALSTTGEIPFYGKAI	348

```
Qy 316 PLEVIKGRHLIFCHSKKKCDLAAKLVAGINAVAYRGLDVSIVPTSGDVVVVATDAL 375
Db 349 PLEVIKGRHLIFCHSKKKCDLAAKLVAGINAVAYRGLDVSIVPTSGDVVVVATDAL 408
Qy 376 MTGYTGDFDSVIDCNTC----- 392
Db 409 MTGYTGDFDSVIDCNTCWTQTVDLSLPTFTIETITLPQDAVSRTQRRGTRGKPGIYR 468
Qy 393 ----- 392
Db 469 FVAPGERPSGMFDSVLCYDAGCAWVELTPAETTVRLRAYMTNPGLPVCQDHLFWE 528
Qy 393 ----- 392
Db 529 VFTGLTHIDAHFLSQTKSGENLPYLWAYQATVCARAQAPPSPWDQMMKCLIRLKP 588
Qy 393 ----- 392
Db 589 PTPLLYRLGAVONEITLTHPTVKYIMTMSADLEVTVTWLVGVGLAALAAAYCLSTGCV 648
Qy 393 -----ACSGKPAIIPREVLYREFDEMECSOHLPIYIEQGMMLAEQFKQKALGL-----S 442
Db 649 VIVGRVLSGKPAIIPREVLYREFDEMECSOHLPIYIEQGMMLAEQFKQKALGLQTAS 708
Qy 443 RGGKPAIVDPKEVLYQQVD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLDNDQVV 497
Db 709 RQAE-VIAPAVQTNWQKLETFWAKHMMNFISIQIYLAGLSLTPG--NPAIASLMAFTA 765
Qy 498 VTP---DKEILYE-----AFDEMEECASKAALIEGORMAELKSKIQ 537
Db 766 TSPUTTSQTLNFGWAAQAALAPGAATFVAGLAGAAGVGLGKVLIDILAGVGA 825
Qy 538 GLLG-----ILRRHVGEGEAGVQWMN 558
Db 826 GVAGALVAFKMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGEGEAGVQWMN 885
Qy 559 RLIAFASRGNHVSPTHVPSRRRFAQALPWARPDYNPLVETWKKPDYPPVPHGRSS 618
Db 886 RLIAFASRGNHVS-----GNSS 903
Qy 619 RRFAQALPVWARPDYNPLVETWKKPDYPPVPHGRKTKRNTNRRPQDVKPPGGQIVGG 678
Db 904 T-----NP-----KPO-----KKNKNTNRRPQDVKPPGGQIVGG 934
Qy 679 VYLLPRRGPRLGLVLAIRKTS-----PIPKARRPEGRRTWAQPGYPWPLYGNK----- 724
Db 935 VYLLPRRGPRLGLVLAIRKTSERSQPRGRQPIPKARRPEGRRTWAQPGYPWPLYNCGCGW 994
Qy 725 -----DRSTGKSGKPGYPWPKTKRN 747
Db 995 AGWLLSPRGRSPSWGPTD---PRRRSRN 1019

RESULT 9
AAE22050
ID AAE22050 standard; protein; 1021 AA.
AC
XX
XX AAE22050;
XX
DT 16-JUL-2002 (first entry)
XX
DE PSOD/c200/core expression plasmid protein.
XX
KW Hepatitis C virus; HCV; antigen; C domain; polyprotein; NS3 domain;
KW NS4 domain; S domain; NS5 domain; PSOD/c200/core plasmid.
XX
OS Hepatitis C virus.
OS Unidentified.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT Region 1..154
FT /note= "hsod"
```

```
FT Region 155..159
FT /note= "Linker region"
FT Region 160..189
FT /note= "HCV c200"
FT Region 900..902
FT /note= "Linker region"
FT Region 903..1021
FT /note= "HCV c22"
XX
XX US6312889-B1.
XX
XX 06-NOV-2001.
XX
XX 12-MAY-1995; 95US-00440549.
XX
XX 04-APR-1990; 90US-00504352.
XX
XX 07-JUL-1992; 92US-00910760.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
XX
XX WPI; 2002-040268/05.
XX
XX N-PSDB; AAD35044.
XX
XX Combination of hepatitis C viral (HCV) antigens, useful in improved
XX immunosassay for detecting HCV antibodies.
XX
XX Example 6; Fig 4; 58pp; English.
XX
XX The invention relates to combination of hepatitis C viral (HCV) antigens
XX that have a broader range of immunological reactivity than any single HCV
XX antigen. The combinations consist of an antigen from the C domain of the
XX HCV polyprotein, and at least one additional HCV antigen from either the
XX NS3 domain, the NS4 domain, the S domain, or the NS5 domain and are in
XX the form of fusion protein, a simple physical mixture, or the individual
XX antigens commonly bound to a solid matrix. The combinations of antigens
XX provides broad range immunoassays for anti-HCV antibodies. The invention
XX therefore provides a method for detecting antibodies to HCV in a mammal
XX suspected of containing such antibodies. The present sequence is a
XX CC protein encoded by pSOD/c200/core expression plasmid DNA containing HCV
XX CC coding sequence
XX
XX Sequence 1021 AA;
XX
XX Query Match 49.9%; Score 2222; DB 5; Length 1021;
XX Best Local Similarity 46.2%; Pred. No. 4.1e-150;
XX Matches 512; Conservative 35; Mismatches 111; Indels 450; Gaps 22;
Qy 1 MATKAVCVLKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGTCTS 60
Db 1 MATKAVCVLKGDPVQGIINPEQKESNGPVKVMGSIKGLTEGLHGFHVEFGDNTAGTCTS 60
Qy 61 AGPHFNPLSTGNCNCSYIPGHIITGHRMAWKLGSARTTSG-----FVSL----- 104
Db 61 AGPHFNPLSRK-----HGGPKDBERHVGDLGNVTADKGVADVSDVIESVLSGDHCII 114
Qy 105 -----FAPGAKQNETHTVTGGAARTTSLTSLFSPGASQNIQ----- 141
Db 115 GRTLNVHEKADDLCKGGNEESTK-TGNAGSLACGVI-----GIAQNLEFGAVDFIPVEN 168
Qy 142 LITS-----TDNSSPPVPOSFOVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNSVAA 195
Db 169 LETTMRSPVFTDNSSPPVPOSFOVAHLHAPTGSKSTKVPAAAYAAQYKVLVLPNSVAA 228
Qy 196 TLGFGAYMSKAHGIDPNIRTVRTITTSPTTYTYGKFLADGGCGGAYDIICDECHS 255
Db 229 TLGFGAYMSKAHGIDPNIRTVRTITTSPTTYTYGKFLADGGCGGAYDIICDECHS 288
Qy 256 TDATSIILGIGTVLDOAETAGARLVVLTATATPPGVTVPHPNIEEVALSTTGEIPFYGKAI 315
Db 289 TDATSIILGIGTVLDOAETAGARLVVLTATATPPGVTVPHPNIEEVALSTTGEIPFYGKAI 348
```

QY	316	PLEVIKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	375
Db	349	PLEVIKGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDAL	408
QY	376	MTGYTGDFSDVCNTC-----	392
Db	409	MTGYTGDFSDVCNTCVTQTVDFTFTIETITLPQDAVSRTQRRGTRGKPGIYR	468
QY	393	-----	392
Db	469	FVAPGRPSGMFSSVLCYDAGCAWYELTPAETTVRLRAYMNTGLPVCQDHLSEWEG	528
QY	393	-----	392
Db	529	VFTGLTHIDAHFLSQTKSGENLPYLVAQYTCARAQAAPPSPDQMKCLIRLKP TLHG	588
QY	393	-----	392
Db	589	PTPLLYRLGAVQNEITLTPVTKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTGCV	648
QY	393	-----ACSGKPAIIDREVLYREFDEMECSOHLPIYEOGMLAQFKOKALGL---S	442
Db	649	VIVGRVLSGKPAIIDREVLYREFDEMECSOHLPIYEOGMLAQFKOKALGLLOTAS	708
QY	443	RGKPAIVDPKVELYQOYD-----EMECSSAAPIYIEQAQVIAHOFKEKVLGLINDQVV	497
Db	709	QRAE-VIAPAVQTNQKLETFWAKHMWNFTSGIQYLAGLSTLPG--NPATASLMAFTAAV	765
QY	498	VTP-----DKEILYE-----AFDEMECSAALIEGQRMALSKSIQ	537
Db	766	TSPLTTSQTLILNIGWVAALAAPGAATAFVAGLAGAAGSVGLGLVILIDILAGYGA	825
QY	538	GLAG-----ILRRHVGPGGAVQWMN	558
Db	826	GVAGALVAFKIMGEVSTEDLVNLLPAILSPGALVGVVCAALLRRHVGPGGAVQWMN	885
QY	559	RLIAFASRGNHVSPTHVPSRSRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRSS	618
Db	886	RLIAFASRGNHVS-----GNSS	903
QY	619	RRFAQALPVMARPDYNPLVETWKKPDYEPVPHGRKTKNTNRRPDQVKFPGGGQIVGG	678
Db	904	T-----NP-----KQO-----KKNKENTNRRPDQVKFPGGGQIVGG	934
QY	679	VYLLPRGRGLVLAIRKTS-----PIKARRPEGRTWAOQGYPMPLYGNK----	724
Db	935	VYLLPRGRGLVLAIRKTSERSQPRGRQPIPKARRPEGRTWAOQGYPMPLYGNKCGW	994
QY	725	-----DRRSTGKSGKPGYPMPRKTKRN	747
Db	995	AGWLLSPRGRSPSWGPTD---PRRSRN	1019
RESULT 10			
AAP92041			
ID	AAP92041 standard; protein; 1766 AA.		
XX			
AC	AAP92041;		
XX			
DT	09-SEP-2004 (revised)		
DT	25-MAR-2003 (revised)		
DT	02-MAR-1990 (first entry)		
DE	Hepatitis C virus (HCV) protein of cDNA clone inserts.		
XX			
KW	Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).		
XX			
OS	Hepatitis C virus.		
XX	Unidentified.		
XX	EP318216-A.		
PN			
XX	31-MAY-1989.		
PD			
XX	18-NOV-1988;	88EP-00310922.	
XX	18-NOV-1987;	87US-00122714.	
PR	30-DEC-1987;	87US-00139886.	
PR	26-FEB-1988;	88US-00161072.	
PR	06-MAY-1988;	88US-00191263.	
PR	26-OCT-1988;	88US-00263584.	
PR	14-NOV-1988;	88US-00271450.	
XX	(CHIR ) CHIRON CORP.		
PA	(CHIR ) CHIRON CORP.		
PI	Houghton M, Choo QL, Kuo G;		
XX	WPI; 1989-159274/22.		
DR	N-PSDB; AAN92097.		
XX	Purified hepatitis C virus - and associated nucleic acids and polypeptide(s).		
PS	Claim 13; Fig 26-1, 26-2, 26-3, 26-4, 26-5, 26-6; 139pp; English.		
XX	It is the sequence encoded in the open reading frame of hepatitis C virus cDNA inserts in clones 141, 11b, 7f, 7e, 8h, 33c, 40b, 37b, 35, 36, 81, 32, 33b, 25c, 14c, 8f, a33f, 33g and 39c. It is antigenic and could be used in immunosay reagents and vaccines and to generate antibodies useful in diagnosis and passive immunotherapy for HCV infection/non-A, non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PI field.)		
CC	Revised record issued on 09-SEP-2004 : Correction to DE line		
XX	Sequence 1766 AA;		
SQ	Query Match 36.5%; Score 1627.5; DB 1; Length 1766; Best Local Similarity 36.6%; Pred. No. 3.7e-107; Mismatches 81; Indels 581; Gaps 15; Matches 398; Conservative 26;		
QY	146	TDNSSPPVPOSPQVLAHLHAPTGSKSTKVPAAYAAQGVKVLNPSVAATLGFAYMSK	205
Db	495	TDNSSPPVPOSPQVLAHLHAPTGSKSTKVPAAYAAQGVKVLNPSVAATLGFAYMSK	554
QY	206	AHGIDPNIRTVRTITTTGSPITVSTYTGKFLADGCGSGAYDIIICDECHSTDATSILGIG	265
Db	555	AHGIDPNIRTVRTITTTGSPITVSTYTGKFLADGCGSGAYDIIICDECHSTDATSILGIG	614
QY	266	TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYKKAIPLEVIKGRH	325
Db	615	TVLDQAEATAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYKKAIPLEVIKGRH	674
QY	326	LIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS	385
Db	675	LIFCHSKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFDS	734
QY	386	VDCNTC-----	392
Db	735	VDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPGRPSG	794
QY	393	-----	392
Db	795	MPDSSVLCYDEGCWYELTPAETTVRLRAYMNTGLPVCQDHLSEWEGVFTGLTHIDA	854
QY	393	-----	392
Db	855	HFLSQTKSGENLPYLVAQYTCARAQAAPPSPDQMKCLIRLKP TLHGPTPLLYRLGA	914
QY	393	-----ACSG	396
Db	915	VQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG	974
QY	397	KPAIIDREVLYREFDEMECSOHLPIYEOGMLAQFKOKALGL-----SRGKPAIVPD	452

```
Db 975 KPAIIPREVLYREFDEMECSQHLPIEQOMMLAEQFKKALGLLOTASQAE-VIAPA 1033
Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1034 VQTNWQKLETFWAKHWNFIISGQYLAGLSLPG--NPAIASLMAFTAAVTSPLTTSTQL 1091
Qy 505 LYE-----AFDEMECASKAALIEGQRMKSKIOGLLG-----541
Db 1092 LFNILGGWVAQAALAPGAATAFVAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFK 1151
Qy 542 -----TLRRHVGPGEAVOMNRLIAFASRGN 568
Db 1152 IMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVOMNRLIAFASRGN 1211
Qy 569 HVSPTHYVPS-----578
Db 1212 HVSPTHYVPSDAAARVTAISSLTVTQLRLRLHQLWISSECTTPCSGSLRDIWDWICEV 1271
Qy 579 -----578
Db 1272 LSDFKTLWAKMLPQLPGIPFVSCQGYGVWRVDMHTRCHGCAEITGHVKNGTMRIV 1331
Qy 579 -----578
Db 1332 GPTRCRNMWGTFFPINAYTTGCTPLPAPNYTFALMRVSAEYVEIRQVPSPEFTLDG 1391
Qy 579 -RSRFA-----QALPVWARPDPY-----595
Db 1392 VRLHRFAPPCKPLREEVSFRVGLHEYPVGSQCEPEPDVAVLTSMLTDPSHITAAG 1451
Qy 596 -----NPP-----LVET---WKK-----605
Db 1452 RRLARGSPPSVASSASQSLKATCTANHDSFDAELIEANLLWRQEMGNITRVESE 1511
Qy 606 -----PDYEPVVG-----RSRRFAQALPVWARPDPYVETWKKPDYE 647
Db 1512 NKKVILSFDPLVAEDEEREISVPAELIRKSRFAQALPVWARPDPYVETWKKPDYE 1571
Qy 648 PPVVG 653
Db 1572 PPVVG 1577

RESULT 11
ID AAP90164 standard; protein; 2261 AA.
XX AAP90164;
AC AAP90164;
DT 25-MAR-2003 (revised)
DT 01-NOV-1989 (first entry)
XX
DE Peptide encoded by composite hepatitis C virus cDNA.
XX
KW Hepatitis C virus; clone 12f; clone 15e; probe; vaccine.
XX
OS Pan troglodytes.
XX
PN GB2212511-A.
XX
PD 26-JUL-1989.
XX
PF 18-NOV-1988; 88GB-00027024.
XX
PR 18-NOV-1987; 87US-00122714.
PR 30-DEC-1987; 87US-00139886.
PR 26-FEB-1988; 88US-00161072.
PR 26-OCT-1988; 88US-00263584.
XX
PA (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo QL, Kuo G;
XX
```

```
DR WPI; 1989-215054/30.
XX N-PSDB; AAN90331.
XX
PT Hepatitis C virus gene - used for prodn. of polynucleotide probes
PT polypeptide(s) and antibodies for diagnosis, prevention and treatment of
PT infection.
XX
PS Disclosure; Fig 32; 30pp; English.
XX
CC The sequence is the peptide encoded by the composite hepatitis C virus
CC (HCV) cDNA of AAN90331. The polypeptides are used to diagnose HCV-induced
CC NANBH, to raise antibodies for immunoassay or treatment, or to produce
CC vaccines. (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 2261 AA;
Query Match 36.5%; Score 1624.5; DB 1; Length 2261;
Best Local Similarity 36.1%; Pred. No. 8.4e-107;
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;
Qy 146 TNSPPVPVPSQFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db 586 TNSPPVPVPSQFQVAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSK 645
Qy 206 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSDTATSILGIG 265
Db 646 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSDTATSILGIG 705
Qy 266 TVLDOATAGARLVVLTATATPPGSVTPHPNIEVALSTTGEIPFYKAIPLVIKGRH 325
Db 706 TVLDOATAGARLVVLTATATPPGSVTPHPNIEVALSTTGEIPFYKAIPLVIKGRH 765
Qy 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTYGDPDS 385
Db 766 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTYGDPDS 825
Qy 386 VIDNCTC-----392
Db 826 VIDNCTCVTQTVDFSLDPTFTIETILPQDAVSTORRGTGRGKPGIYRFVAGERPSSG 885
Qy 393 -----392
Db 886 MFDSSVLCEYDAGCANVELTPAETTVRLRAYMNTPGLPVQDHLFEWGVFTGLTHIDA 945
Qy 393 -----392
Db 946 HFLSOTKQSGENLPYVAYQATVCARQAQPPSPDQMKLIRLKLPTLHGPTLLYRLGA 1005
Qy 393 -----ACSG 396
Db 1006 VQNEITLTHPVTKYIMTCSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVLUSG 1065
Qy 397 KPAIIPREVLYREFDEMECSQHLPIEQOMMLAEQFKKALGL-----SRGKPAIYVD 452
Db 1066 KPAIIPREVLYREFDEMECSQHLPIEQOMMLAEQFKKALGLLOTASQAE-VIAPA 1124
Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1125 VQTNWQKLETFWAKHWNFIISGQYLAGLSLPG--NPAIASLMAFTAAVTSPLTTSTQL 1182
Qy 505 LYE-----AFDEMECASKAALIEGQRMKSKIOGLLG-----541
Db 1183 LFNILGGWVAQAALAPGAATAFVAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFK 1242
Qy 542 -----TLRRHVGPGEAVOMNRLIAFASRGN 568
Db 1243 IMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVOMNRLIAFASRGN 1302
Qy 569 HVSPTHYVPS-----578
Db 1303 HVSPTHYVPSDAAARVTAISSLTVTQLRLRLHQLWISSECTTPCSGSLRDIWDWICEV 1362
Qy 579 -----578
```

Db 1363 LSDFKTLKAKLMPQLPGIPFVSCQRYKGVWRVGDGIMHTRCHCGAEITGHVKNGTMRIV 1422  
QY 579 ----- 578  
Db 1423 GPRTCRNWMSGTFPINAYTTGPTCLPAPNYTFALMRVSAEYVEIRQVGFHYVTGMTT 1482  
QY 579 -----RSRRA-----QALPVMWARP 594  
Db 1483 DNLKPCQVSPFEFFTELDGVLRLHRAAPPCKLLREVSFRVLHHEYVPGSQLPCEPEPD 1542  
QY 595 Y-----NPP-----LV 600  
Db 1543 VAVLTSMLTDPSSHITAEAGRRRLARGSPPSVASSASQLSAPSKATCTANHDSPOAELI 1602  
QY 601 ET---WKK-----PDYEPVVHG-----RSRRAQALPV 627  
Db 1603 EANLLWRQEMGNITRVESENKVVILDSFDPLVAEEDEREISVPAEILRKSRRAQALPV 1662  
QY 628 WARPDPNPLVETWKKPDYEPVVHG 653  
Db 1663 WARPDPNPLVETWKKPDYEPVVHG 1688

## RESULT 12

AAP92050  
ID AAP92050 standard; protein; 2436 AA.

XX AAP92050;

XX 09-SEP-2004 (revised)

DT 25-MAR-2003 (revised)

DT 02-MAR-1990 (first entry)

XX HCV protein of the cDNA inserts in clones K9-1 through 15e.

DE Hepatitis C virus (HCV); non-A, non-B hepatitis (HANBH).

XX Hepatitis C virus.

XX Hepatitis C virus.

OS Unidentified.

XX EP318216-A.

XX 31-MAY-1989.

XX 18-NOV-1988; 88EP-00310922.

XX 18-NOV-1987; 87US-00122714.

PR 30-DEC-1987; 87US-00139886.

PR 26-FEB-1988; 88US-00161072.

PR 06-MAY-1988; 88US-00191263.

PR 26-OCT-1988; 88US-00263584.

PR 14-NOV-1988; 88US-00271450.

XX (CHIR ) CHIRON CORP.

PA (CHIR ) CHIRON CORP.

XX Houghton M, Choo Q, Kuo G;

XX WPI; 1989-159274/22.

DR N-PSDB; AAN92106.

XX Purified hepatitis C virus - and associated nucleic acids and

PT polypeptides).

XX Claim 13; Fig 47-1-47-8; 139pp; English.

XX It is the sequence encoded in the open reading frame of hepatitis C virus

CC (HCV) cDNA inserts in clones K9-1 through 15e. It is antigenic and could

CC be used in immunoassay reagents and vaccines and to generate antibodies

CC useful in diagnosis and passive immunotherapy for HCV infection/non-A,

CC non-B hepatitis. (Updated on 25-MAR-2003 to correct PR field.) (Updated

CC on 25-MAR-2003 to correct PI field.)

CC Revised record issued on 09-SEP-2004 : Correction to DE line  
XX  
SQ Sequence 2436 AA;  
Query Match 36.5%; Score 1624.5; DB 1; Length 2436;  
Best Local Similarity 36.1%; Pred. No. 9.3e-107;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;  
QY 146 TONSSPPVPOQSOVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205  
Db 761 TONSSPPVPOQSOVAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 820  
QY 206 AHGIDNIRITGRTITGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDATSILGIG 265  
Db 821 AHGIDNIRITGRTITGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDATSILGIG 880  
QY 266 TVLDQAEATAGARLVLATATPPGSVTVPHNTEVALSTTGEIPFYGKAIPLEVIKGRH 325  
Db 881 TVLDQAEATAGARLVLATATPPGSVTVPHNTEVALSTTGEIPFYGKAIPLEVIKGRH 940  
QY 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDGDS 385  
Db 941 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDGDS 1000  
QY 386 VIDCNTC----- 392  
Db 1001 VIDCNTCVTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGRKPGIYRFVAPGERPSG 1060  
QY 393 ----- 392  
Db 1061 MFDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFEGVFTGLTHIDA 1120  
QY 393 ----- 392  
Db 1121 HFLSOTKQGENLPYLVAQVATVCARAQAPPSPSDQMWKCLIRLKLPTLHPTPLLYLGA 1180  
QY 393 -----ACSG 396  
Db 1181 VQNEITLTHPTVKYIMTCHSADILEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVVLSG 1240  
QY 397 KPAPIIDREVLYREFDEMEBECQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452  
Db 1241 KPAPIIDREVLYREFDEMEBECQHLPIYIEQGMMLAEQFKQKALGLLQATASRQAE-VIAPA 1299  
QY 453 KEVLYQQYD-----EMEBECQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504  
Db 1300 VQTNWQKLETFWAKHMWNFISGQYLAGLSTLPG--NPATIASLMAFTAATVTSPLTTSQTL 1357  
QY 505 LYE-----AFDEMEBECASKAALIEBQORMAEMLKSKIOGLLG----- 541  
Db 1358 LFNILGWWAAQLAAPGAATAFVAGLAGAAGAAIGSVGLKVLIDILAGYAGAGVAGALVAFK 1417  
QY 542 -----ILRRHVCPGEGAVQMMNRLTAFASRGN 568  
Db 1418 IMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVCPGEGAVQMMNRLTAFASRGN 1477  
QY 569 HVSPTHYVRS----- 578  
Db 1478 HVSPTHYVRS----- 578  
QY 579 ----- 578  
Db 1538 LSDFKTLKAKLMPQLPGIPFVSCQRYKGVWRVGDGIMHTRCHCGAEITGHVKNGTMRIV 1597  
QY 579 -----RSRRA----- 578  
Db 1598 GPRTCRNWMSGTFPINAYTTGPTCLPAPNYTFALMRVSAEYVEYRQVGFHYVAGMIT 1657  
QY 579 -----RSRRA-----QALPVMWARP 594  
Db 1658 DNLKPCQVSPFEFFTELDGVLRLHRAAPPCKLLREVSFRVLHHEYVPGSQLPCEPEPD 1717



Qy	595 Y-----NPP-----	LV 600
Db	1718 VAVLTSMLTDPSSHITAEAGRLRAGSPSVASSASQLSAPSLKATCTANHDSFDAELI	1777
Qy	601 ET---WKK-----PDYPPVHVG-----RSSRFAQALPV	627
Db	1778 EANLLWQEGMGNITRVESENKVVILDSFDPLVAEEDEREISVPAEILRKSRFAQALPV	1837
Qy	628 WARDYNPPLVETWKKPDYPPVHVG 653	
Db	1838 WARDYNPPLVETWKKPDYPPVHVG 1863	
RESULT 13		
AA	AAP90288	
ID	AAP90288 standard; protein; 2436 AA.	
AC	AAP90288;	
XX		
DT	25-MAR-2003 (revised)	
DT	19-JUL-2001 (revised)	
DT	01-NOV-1989 (first entry)	
XX	Peptide encoded by composite hepatitis C cDNA.	
DE	Hepatitis C virus; clone 15e; clone k9-1; probe; vaccine.	
XX		
KW	Pan troglodytes.	
XX		
OS		
XX	GB2212511-A.	
PN		
XX	26-JUL-1989.	
PD		
XX	18-NOV-1988; 88GB-00027024.	
PF		
XX	18-NOV-1987; 87US-00122714.	
PR	30-DEC-1987; 87US-00139886.	
PR	26-FEB-1988; 88US-00161072.	
PR	26-OCT-1988; 88US-00263584.	
XX	(CHIR ) CHIRON CORP.	
PA		
XX	Houghton M, Choo QL, Kuo G;	
PI		
XX	WPI; 1989-215054/30.	
DR	N-PSDB; AAN90336.	
XX	Hepatitis C virus gene - used for prodn. of polynucleotide probes	
PT	polypeptide(s) and antibodies for diagnosis, prevention and treatment of	
PT	infection.	
XX	Disclosure; Fig 47-1 to 47-8; 30pp; English.	
PS		
XX	The sequence is the peptide encoded by the composite hepatitis C virus	
CC	(HCV) cDNA of AAN90336. The polypeptides are used to diagnose HCV-induced	
CC	NANBH, to raise antibodies for immunoassay or treatment, or to produce	
CC	vaccines. (N.B. This record was resubmitted to correct errors in the	
CC	sequence.) (Updated on 25-MAR-2003 to correct PR field.)	
XX		
XX	Sequence 2436 AA;	
Qy	Query Match 36.5%; Score 1624.5; DB 1; Length 2436;	
Db	Best Local Similarity 36.1%; Pred. No. 9.3e-107;	
XX	Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;	
Qy	146 TDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAAAYAAQGYKVLVNLNPSVAATLGFAYMSK 205	
Db	761 TDNSSPPVVPQSFQVAHLHAPTGSGKSTKVPAAAYAAQGYKVLVNLNPSVAATLGFAYMSK 820	
Qy	206 AHGIDPNRTGVRTITGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATSIIGIG 265	
Db	821 AHGIDPNRTGVRTITGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATSIIGIG 880	

Qy	266 TVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYKAIPLVIKGRH	325
Db	881 TVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYKAIPLVIKGRH	940
Qy	326 LIFCHSKKKCDELAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDS	385
Db	941 LIFCHSKKKCDELAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDS	1000
Qy	386 VIDCNTC-----	392
Db	1001 VIDCNTCTQTVDPSLDPTFTTITLPODAVSTQRRGRTGRGKPGIYRFVAPGERPSG	1060
Qy	393 -----	392
Db	1061 MFDSSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLTHIDA	1120
Qy	393 -----	392
Db	1121 HFLSQTKSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKLPTLHGPTLLYRLGA	1180
Qy	393 -----	392
Db	1181 VQNEITLTHPVTKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG	1240
Qy	397 KPATIPDREVLYREFDEMESCQHLPIYIEQGMMLAEQKOKALGL-----SRGGKPAIVPD	452
Db	1241 KPATIPDREVLYREFDEMESCQHLPIYIEQGMMLAEQKOKALGLLQATASQAB-VIAPA	1299
Qy	453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKXI	504
Db	1300 VQTNWQKLETFWAKGMNFIQIYLAGLSLPG--NPAITASLMAFTAATVSPLTTSQTL	1357
Qy	505 LYE-----AFDEMEECSKAALIEGQMAEMLKSKIOGLG-----	541
Db	1358 LFNILGGWAAQLAAPGAATAFVGAGLAGAIGSVGLKVLIDILAGYGAGVAGALVAFK	1417
Qy	542 -----TLRRHVGPGEAVOMNRLIAFASRGN	568
Db	1418 IMSGEVSTEDLVNLLPALISPGALVGVVCAAILRRHVGPGEAVOMNRLIAFASRGN	1477
Qy	569 HVSPTHVVP-----	578
Db	1478 HVSPTHVVPESDAAARVTAISSLVTQLLRLHQMISSECTTFCSGSWLRDINDWICEV	1537
Qy	579 -----	578
Db	1538 LSDFKTLWAKLMPQLPGIPFVSCQGYGVWRVDGIMHTRCHGAEITGHVKGNTMRIV	1597
Qy	579 -----	578
Db	1598 GPRTCRNMMSGTFPFINAYTTGCTPLPAPNVTFALMRVSAEEYVEIROGVDFHYVTGWT	1657
Qy	579 -----RSRFA-----QALPVWARP	594
Db	1558 DNLKPCQVPSPEFFTELDGVRLHRFAPPCKPLLRREEVSFRVLHXYVPGSQLPCEPDP	1717
Qy	595 Y-----NPP-----LV	600
Db	1718 VAVLTSMLTDPSSHITAEAGRLRAGSPSVASSASQLSAPSLKATCTANHDSFDAELI	1777
Qy	601 ET---WKK-----PDYPPVHVG-----RSSRFAQALPV	627
Db	1778 EANLLWQEGMGNITRVESENKVVILDSFDPLVAEEDEREISVPAEILRKSRFAQALPV	1837
Qy	628 WARDYNPPLVETWKKPDYPPVHVG 653	
Db	1838 WARDYNPPLVETWKKPDYPPVHVG 1863	
RESULT 14		
AA	AAB18540	
ID	AAB18540 standard; protein; 2772 AA.	
XX		

AC AAB18540;  
XX  
DT 15-JAN-2001 (first entry)  
XX Protein encoded by a cDNA compiled Hepatitis C virus cDNA clones.  
DE  
XX Hepatitis C virus; HCV; antisense polynucleotide; polyprotein;  
KW viral infectivity; viral replication.  
XX  
OS Hepatitis C virus.  
XX  
XX EP1034785-A2.  
XX  
XX 13-SEP-2000.  
XX  
XX 16-MAR-1990; 2000EP-00109602.  
XX  
XX 17-MAR-1989; 89US-00325338.  
PR 20-APR-1989; 89US-00341334.  
PR 18-MAY-1989; 89US-00355002.  
PR 16-MAR-1990; 90EP-00302866.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Houghton M, Choo Q, Kuo G;  
PI  
XX  
XX WPI; 2000-566891/53.  
DR N-PSDB; AANV5296.  
XX  
XX Novel composition comprising a hepatitis C virus antisense polynucleotide  
PT which is complementary to or corresponds to a sense strand of the virus  
PT genome, and selectively hybridizes to it.  
XX  
XX Example; Fig 16; 75pp; English.  
XX  
XX The specification describes a pharmaceutical composition which comprises  
CC a hepatitis C virus (HCV) antisense polynucleotide. The HCV is  
CC characterized by a positive stranded RNA genome which has 40% homology at  
CC the polypeptide level to a HCV polypeptide. The antisense polynucleotide  
CC binds to cellular polynucleotides which enhance and/or are required for  
CC viral infectivity, replicative ability or chronicity. The antisense  
CC polynucleotides may also be designed to bind with high specificity, to be  
CC of increased stability, to be stable and to have low toxicity. The  
CC composition also comprises an agent which causes viral RNA to be  
CC inactive. The composition is used for preventing HCV replication in a  
CC system. The present sequence is encoded by a novel HCV cDNA sequence,  
CC which is used in the course of the invention  
XX  
SQ Sequence 2772 AA;  
  
Query Match 36.5%; Score 1624.5; DB 3; Length 2772;  
Best Local Similarity 36.1%; Pred. No. 1.1e-106;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;  
  
QY 146 TNSPPVPPQSFQVLAHLHPTGSGKTKVPAAYAAQYKVLNPSVAATLGFAYMSK 205  
DB 1097 TNSPPVPPQSFQVLAHLHPTGSGKTKVPAAYAAQYKVLNPSVAATLGFAYMSK 1156  
  
QY 206 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 265  
DB 1157 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 1216  
  
QY 266 TVLDQAEATAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKKAIPLEVIKGRH 325  
DB 1217 TVLDQAEATAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKKAIPLEVIKGRH 1276  
  
QY 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 385  
DB 1277 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 1336  
  
QY 386 VIDCNTC----- 392  
DB 1337 VIDCNTCVTQVDFSLDPTFTIETITLPQDAVSRTORRGTRGKGIYRFVAPGRPSG 1396

QY 393 ----- 392  
DB 1397 MFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLBEWGVFTGLTHIDA 1456  
QY 393 ----- 392  
DB 1457 HFLSOTKQSGENLPYLVAQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPLLYRLGA 1516  
QY 393 -----ACSG 396  
DB 1517 VQNEITLTHPVTKYIMTCSADLEVVSTWLVGGVLAALAAAYCLSTGCVVIVGRVLSG 1576  
QY 397 KPAAIIPDREVLYREFDEMEECQHLPIYIQGQWMLAEQFKQKALGL-----SRGKPAIVPD 452  
DB 1577 KPAAIIPDREVLYREFDEMEECQHLPIYIQGQWMLAEQFKQKALGLLQATASQAE-VIAPA 1635  
QY 453 KEVLYQOYD-----EMEECSQAAPYIEQAQVTAHQFKKVKVLGIDNDQVVVTP---DKEI 504  
DB 1636 VQTNQKLETFWAKHWNFIISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSOTL 1693  
QY 505 LYE-----AFDEMEECASKAALIEEGORMAEMLSKIQGLLG----- 541  
DB 1694 LFNILGWAAQLAAPGAATAFVGAGLAGAAGTSGVGLKVLIDILAGYGAGVAGALVAFK 1753  
QY 542 -----ILRRHVGPGEVAVQWMMNRLIAFASRG 568  
DB 1754 IMSGEVPTEDLVNLLPAILSPCALVGVVCAAILRRHVGPGEVAVQWMMNRLIAFASRG 1813  
QY 569 HVSPTHYVPS----- 578  
DB 1814 HVSPTHYVPSDAAARVTAISSLTVTQLRLRHQWISSECTTPCSGSLRDIWDWICEV 1873  
QY 579 ----- 578  
DB 1874 LSDFKTWLAKMLPQLPGIPFVSCQRYGVWRVDMITRCHGCAEITGHVXNGTMRIV 1933  
QY 579 ----- 578  
DB 1934 GPTRCNMWSGTFPINAYTTGCTPLPAPNYTFALWRVSAEEVVEIROVGDHVTGTT 1993  
QY 579 -----RSRRFA-----QALPVHARPD 594  
DB 1994 DNLCPCQVPSPFEFFTELDGVRHLRFAPFCKPLLREEVSFRVGLHBYPVGSQLPCEPEPD 2053  
QY 595 Y-----NPP-----LV 600  
DB 2054 VAVLTSMLTDPSSHITABAGRRRLARGSPPSVASSASQLSAPSLKATCTANHDSPDAL 2113  
QY 601 ET---WKK-----PDYEPYVH-----RSSRFAQALPV 627  
DB 2114 EANLLWRQMGGNITRVESENKVVILDSFDPDLVAEEDEREISVPAEILRKRRAQALPV 2173  
QY 628 WARPDPNPPILVETWKKPDYEPYVH 653  
DB 2174 WARPDPNPPILVETWKKPDYEPYVH 2199  
  
RESULT 15  
ADN35976  
ID ADN35976 standard; protein; 2772 AA.  
XX  
XX ADN35976;  
XX AC  
XX DT 17-JUN-2004 (first entry)  
XX HCV cDNA clone #1 protein.  
XX Antiviral; Vaccine; hepatitis C virus infection; HCV infection.  
XX Hepatitis C virus.  
XX EP1394255-A2.  
PN

```

XX 03-MAR-2004.
XX 16-MAR-1990; 2003EP-00016585.
XX 17-MAR-1989; 89US-00325338.
XX 20-APR-1989; 89US-00341334.
XX 18-MAY-1989; 89US-00350002.
XX 16-MAR-1990; 90EP-00302866.
XX (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
XX WPI; 2004-193149/19.
XX N-PSDB; ADN35977.
XX
XX Novel purified hepatitis C virus polypeptide comprising epitope encoded
XX by hepatitis C virus cDNA, useful as vaccine for treating hepatitis C
XX virus.
XX
XX Example 1; Fig 16; 79pp; English.
XX
XX The present invention relates to hepatitis C virus (HCV) proteins and
XX cDNA sequences. The sequences are useful in immunoassays for detecting
XX antibodies directed against HCV antigen; preparing host cells transformed
XX with a recombinant polynucleotide; screening antiviral agents and
XX determining the effect of antiviral agent in inhibiting viral replication
XX in cell culture system; and developing vaccine for treating HCV
XX infection.
XX
XX Sequence 2772 AA;
XX
XX Query Match 36.5%; Score 1624.5; DB 8; Length 2772;
XX Best Local Similarity 36.1%; Pred. No. 1.1e-106;
XX Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;
XX
XX 146 TDNSSPPVPPQSFQVLAHLAAPTSGKSTKVPAAQAQYKVLNPNVAATLFGAYMSK 205
XX 1097 TDNSSPPVPPQSFQVLAHLAAPTSGKSTKVPAAQAQYKVLNPNVAATLFGAYMSK 1156
XX
XX 206 AHGIDPNIRGTGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 265
XX 1157 AHGIDPNIRGTGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDCHSTDATSIILGIG 1216
XX
XX 266 TVLDQAEATAGARLVLAATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 325
XX 1217 TVLDQAEATAGARLVLAATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLEVIKGRH 1276
XX
XX 326 LIFCHSKKCDLAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 385
XX 1277 LIFCHSKKCDLAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 1336
XX
XX 386 VIDCNTC----- 392
XX 1337 VIDCNTCVTQVDFSLDPTFTIETITLPQDAVSRQRTGRGKPGIYRFVAPGERPSG 1396
XX
XX 393 ----- 392
XX
XX 1397 MFDSSVLCECYDAGCANVELTPARTTVLRAYMNTPGLPVCQDHLPEWEGVFTGLTHIDA 1456
XX 393 ----- 392
XX
XX 1457 HFLSOTKQSGENLPYLVAQATVCARAQAPPPSNDQWKCLIRLKLPTLHGPTLLYRLGA 1516
XX 393 -----ACSG 396
XX
XX 1517 VQNEITLTHPVTKYIMTCSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRVVLSG 1576
XX
XX 397 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKOKALGL-----SRGKGPAIYVPD 452
XX 1577 KPAIIPREVLYREFDEMECSQHLPIYEQGMMLAEQFKOKALGLLQATASRAQAE-VIAPA 1635

```

```

Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTP---DKEI 504
Db 1636 VQTNWQKLETFWAKHMFNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPTTTSQTL 1693
Qy 505 LYE-----AFDEMBECASKAALIBEGQRMAMELKSKIOGLLG----- 541
Db 1694 LFNILGGWAAQAAPCAATAFVAGLAGAAGISVGLGKVLIDILAGYGAGVAGAVAFK 1753
Qy 542 -----ILRRHVGPGGGAVQVMNRLIAFASRGN 568
Db 1754 IMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGGGAVQVMNRLIAFASRGN 1813
Qy 569 HVSPTHYVPS----- 578
Db 1814 HVSPTHYVPSDAAARVAILSLTVTLRLRHQWISSSECTTPCSGSLRDIDWDWICEV 1873
Qy 579 ----- 578
Db 1874 LSDFKTWLKAQLPQPGIPFVSCQRGYKGVWRVDGIMHTRCHCGAEITGHVKNGTMRIV 1933
Qy 579 ----- 578
Db 1934 GPRTCRMMSGTTPINAYTTGPCTPLPAPNYTFALMRVSAEYVEIRQVGDHFHYVTGMTT 1993
Qy 579 -----RSRREA-----QALPVMWARP 594
Db 1994 DNLKCPQVPSBFFTELDGVLRLHFAAPPCKPLLRVEVSPRVGLHVEYVPGSQLPCEPEPD 2053
Qy 595 Y-----NPP-----LV 600
Db 2054 VAVLTSMLTDPSSHITAEAGRRRLARGSPPSVASSASQLSAPSLKATCTANHDSDAELI 2113
Qy 601 ET---WKK-----PDYEPVPHG-----RSRRRPAQALPV 627
Db 2114 EANLLWRQEMGNIITRVESENKVVILDSFDPVLAEEDEBISVPABILRKSRRPAQALPV 2173
Qy 628 WARPDPNPLVETWKKPDYEPVPHG 653
Db 2174 WARPDPNPLVETWKKPDYEPVPHG 2199

```

Search completed: November 7, 2005, 20:10:17  
Job time : 117.776 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:00:21 ; Search time 17.1255 Seconds  
(without alignments)  
4657.604 Million cell updates/sec

Title: US-10-658-782-4  
Perfect score: 4455  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRSTGKSWGKPGYPWP 829

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1624.5	36.5	3011	1 GNWVC3	genome polyprotein
2	1596	35.8	3011	1 S40770	genome polyprotein
3	1559.5	35.0	3011	1 GNWVCH	genome polyprotein
4	1534	34.4	3010	1 S18030	genome polyprotein
5	1532.5	34.4	3010	1 A45573	genome polyprotein
6	1523.5	34.2	3010	1 GNWVTC	genome polyprotein
7	1522.5	34.2	3010	1 GNWVTW	genome polyprotein
8	1519	34.1	3010	1 GNWVCJ	genome polyprotein
9	1429.5	32.1	3033	1 JQ1303	genome polyprotein
10	1422.5	31.9	3033	1 GNWVJ8	genome polyprotein
11	1337	30.0	3014	1 JCS620	genome polyprotein
12	1128.5	25.3	386	2 S68016	ATPase/RNA helicase
13	954.5	21.4	876	2 PC2219	polypeptide - hepa
14	860	19.3	216	2 S21337	genome polyprotein
15	608	13.6	3005	2 T08841	polyprotein - dour
16	586.5	13.2	2970	2 T08839	polyprotein - mara
17	409.5	9.2	492	2 PS0326	polyprotein - hepa
18	390	8.8	513	4 T43640	structural protein
19	387.5	8.7	179	4 T43640	superoxide dismuta
20	385	8.6	118	2 S41346	genome polyprotein
21	385	8.6	441	2 S12707	genome polyprotein
22	382.5	8.6	154	1 DSHUC7	superoxide dismuta
23	381	8.6	369	2 S21471	genome polyprotein
24	381	8.6	513	2 PC1284	genome polyprotein
25	380	8.5	178	2 PS0388	genome polyprotein
26	377	8.5	189	2 S32740	polyprotein - hepa
27	377	8.5	550	2 JH0711	genome polyprotein
28	377	8.5	782	2 S18031	genome polyprotein
29	377	8.5	787	2 PN0677	hypothetical prote

RESULT 1

GNWVC3

genome polyprotein - hepatitis C virus (strain HCV-1)  
N/Contains: capsid protein C, envelope protein M; hepatitis virus (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C/Species: hepatitis C virus  
C/Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004  
C/Accession: A39166; PQ0403; PQ0404  
R/Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coi Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991  
A/Title: Genetic organization and diversity of the hepatitis C virus.  
A/Reference number: A39166; MUID:91172826; PMID:1848704  
A/Accession: A39166  
A/Molecule type: mRNA  
A/Residues: 1-3011 <CHO>  
A/Cross-references: UNIPROT:P26664; GB:M62321; NID:G329873; PIDN:AAA45676.1; PID:G329874  
J.Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L. J. Gen. Virol. 73, 1131-1141, 1992  
A/Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to A/Reference number: PQ0393; MUID:92268871; PMID:1316939  
A/Accession: PQ0403  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CHA>  
A/Cross-references: DDBJ:D10128  
A/Experimental source: isolates E-b16  
A/Accession: PQ0404  
A/Status: preliminary  
A/Molecule type: genomic RNA  
A/Residues: 1577-1633 <CH2>  
A/Experimental source: isolates E-b17  
C/Superfamily: hepatitis C virus genome polyprotein  
F/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural F/116-191/Product: envelope protein M #status predicted <CPC>  
F/192-389/Product: major envelope protein E #status predicted <MEB>  
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F/1007-1615/Product: hepatitis virus #status predicted <NS3>  
F/1230-1237/Region: nucleotide-binding motif A (P-loop)  
F/1312-1317/Region: nucleotide-binding motif B  
F/1316-1319/Region: DEXH motif  
F/1616-1862/Product: nonstructural protein NS4a #status predicted <NAA>  
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NAB>  
F/2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F/196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22.22

Query Match 36.5%; Score 1624.5; DB 1; Length 3011;

Best Local Similarity 36.1%; Pred. No. 1.9e-92;

Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;

QY 146 TNSSPFVVPQSFQVHLHAPGSGKSTKVPAAAGYKVLVNPSTVAATLGFQAYMSK 205

|||||



Qy 393 ----- 392  
Db 1586 LVAYQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPLLYRLGAVQGEVTLTHPVTKYI 1645  
Qy 393 -----ACSGKPAIPDREVLYREF 411  
Db 1646 MTCMSADLEVVTSTWLVGVGLAALAYCLSTGCWIVGRVLSGRPAIPDREVLYREF 1705  
Qy 412 DEMEECSQHLPIYIEQGMMLEAFQKQKALGL-----SRGKPAIPDKEVLYQOYD-----E 462  
Db 1706 DEMEECSQHLPIYIEQGMMLEAFQKQKALGLQATASRAE-VIAPTQVNTWQKLEAFNAKH 1764  
Qy 463 MECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVTTP-----DKELIYE----- 507  
Db 1765 MNFISGIQYLAGLSTLPG-----NPAIASLMFAFTAAVTSPLTTSQTLLENILGWVAAQLAA 1822  
Qy 508 -----AFDEMEECASKAALIEBEGORMAEMLSKIQGLLG----- 541  
Db 1823 PGAATAFVSGSLAGAVSGVGLGRVLVDILAGYGAGVAGALVAPKIMSGELPSTEDLVNL 1882  
Qy 542 -----ILRRHVGPGEVAVQVMMNRLIAFASRGNHVSPTHYVPS----- 578  
Db 1883 LPAILSPGALVGVVCAAILRRHVGPGEVAVQVMMNRLIAFASRGNHVSPTHYVPSDDAA 1942  
Qy 579 ----- 578  
Db 1943 RVTAILSLVTQLRLRLHOWLSSESTTPPCSGMWLRDIWDWICEVLSDFKTLWTKLMPH 2002  
Qy 579 ----- 578  
Db 2003 LPIGIPFVSCQHYGVWRGDMHTRCHGAEITHGVKNQGMWIRVGPKTCRMMSGTPTPI 2062  
Qy 579 ----- 578  
Db 2063 NAYTGPCTPLPAPNYTFALMRVSAEYVEIRRVGDFHYVTGWTDTNLKCPQVPSPEFF 2122  
Qy 579 -----RSRRFA-----QALPVWARPDY----- 595  
Db 2123 TELDGVLRHRTAPPCKPCLLREBVSFVGLHDYVPGSQLPCBPEDPAVLTSMLTDPSSHIT 2182  
Qy 596 -----NPP-----LVET-----WKK----- 605  
Db 2183 AAAAGRLARGSPFSEASSASQASAPSLKATCTINHSDPAELIEANLLWRQENGGNIT 2242  
Qy 606 -----PDYEPVVVHG-----RSSRFAQALPVWARPDPYPLVETWK 642  
Db 2243 RVESENKVVILDSFDPLVAEEDEREISVPAELIRKSRFTQALPIWARPDPYPLVETWK 2302  
Qy 643 KPDYEPVVVHG 653  
Db 2303 KPNYEPVVVHG 2313

RESULT 3  
GNMVCH  
genome polyprotein - hepatitis C virus (strain H)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H)  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
A:Note: host Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
R:Accession: A36814; A41546  
R:Inchoupe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
submitted to GenBank, July 1992  
A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: C  
A:Reference number: A36814  
A:Accession: A36814  
A:Molecule type: genomic RNA  
A:Residues: 1-3011 <INC>  
A:Cross-references: UNIPROT:P27958; GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738  
R:Inchoupe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar

A:Reference number: A41546; MUID:92052256; PMID:1658800  
A:Contents: annotation  
A:Note: neither amino acid nor nucleotide sequence is given  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura  
F:1-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <BPM>  
F:192-389/Product: major envelope protein E #status predicted <MSE>  
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>  
F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
F:1312-1317/Region: nucleotide-binding motif B  
F:1316-1319/Region: DEXH motif  
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F:196, 209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 23

Query Match 35.0%; Score 1559.5; DB 1; Length 3011;  
Best Local Similarity 34.9%; Pred. No. 2.2e-88;  
Matches 388; Conservative 32; Mismatches 80; Indels 611; Gaps 15;  
Qy 146 TDNSPPVPOSFOVAHLHAPTGSKSTKPAAYAAQYKVLNPNPSVAATLGFQAYMSK 205  
Db 1211 TDNSPPAVPOSFOVAHLHAPTGSKSTKPAAYAAQYKVLNPNPSVAATLGFQAYMSK 1270  
Qy 206 AHGIDPNIRTVRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSILGIG 265  
Db 1271 AHGVDPNIRTVRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSISGIG 1330  
Qy 266 TVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTTGEIPEYKAIPLEVKGRGH 325  
Db 1331 TVLDOAETAGARLVVLTATATPPGSVTVPHNIEEVALSTTGEIPEYKAIPLEVKGRGH 1390  
Qy 326 LIFCHSKKKDELAAKVALGINAVYRGLDVSVPTSGDVVVVATDALTMTGYTGDFDS 385  
Db 1391 LIFCHSKKKDELAAKVALGINAVYRGLDVSVPTSGDVVVVATDALTMTGYTGDFDS 1450  
Qy 386 VIDNCNTC----- 392  
Db 1451 VIDNCNTVQTQVDFSLDPTFTTITLTPQDAVSRQRRRTGRGKPGIYRFVAPGERPSG 1510  
Qy 393 ----- 392  
Db 1511 MFDSSVLCYDAGCAMPYELTPAETTVRLRAYMNTPGLPVCDHILGFEGVETGLTHIDA 1570  
Qy 393 ----- 392  
Db 1571 HFLSQTQKSGENFPYLVAYQATVCARAQAPPSWDQMKCLIRLKLPTLHGPTPLLYRLGA 1630  
Qy 393 -----ACSG 396  
Db 1631 VQNEVTLTHPTIKYIMTQMSADLEVVTSTWLVGVGLAALAYCLSTGCWIVGRVLSG 1690  
Qy 397 KPAIIPDREVLYREFDEMEECQHLPIYIEQGMMLEAFQKQKALGL-----SRGKPAIPVD 452  
Db 1691 KPAIIPDREVLYREFDEMEECQHLPIYIEQGMMLEAFQKQKALGLQATASRAE-VIAPA 1749  
Qy 453 KEVLYQOYDEMEECQAAPIYIEQAQVIAHQFKEKVLGLINDQV-----VT 499  
Db 1750 VQTNWQKL-EVPWAKHMFISGIQYLAG-----LSTLPGNPATASLMFAFTAAVTSPLT 1802  
Qy 500 PDKELIYE-----AFDEMEECASKAALIEBEGORMAEMLSKIQGLLG- 541  
Db 1803 TGQTLLENILGWVAAQLAAPGAATAFVGCAGLAGAALDSVGLKVLVDILAGYGAGVAGA 1862  
Qy 542 -----ILRRHVGPGEVAVQVMMNRLIAF 563  
Db 1863 LVAFKIMSGEVPSTEDLVNLLPAILSPGALAVGVVFAFASILRRRRVGPGEVAVQVMMNRLIAF 1922  
Qy 564 ASRGNHVSPTHYVPS----- 578

Db 1923 ASRGNHVSPTHYVPESDAAARVTAILSSLTVTQLRLHQMISSECTTPCSGSLWRDIWD 1982  
 QY 579 ----- 578  
 Db 1983 WICEVLSDFKTLWAKLMPQLPGIPFVSCORGYGVRGDCIMHTRCHGCAEITGHVNG 2042  
 QY 579 ----- 578  
 Db 2043 TMRIVGPRCKNMWGTFFINAYTTGPTPLPAPNYKFAKLRVSAEYVEIRRVGDPHYV 2102  
 QY 579 -----RSRRFA-----QALPV 589  
 Db 2103 SGMNTDNLKCPQIQSPPEFFTELDGVLRLHFPAPCKPLLRREVSFRVGLHEYPVGSQPLC 2162  
 QY 590 WARPDPY-----NRP----- 598  
 Db 2163 EPEPDVAULTSMLTDPSSHITAAGRLARGSPSPMASSASQSLASPSLKATCTANHDSP 2222  
 QY 599 ---LVET---WKK-----PDYEPVPHG-----RSSRFA 622  
 Db 2223 DAELIEANLLWRQBMGGNITRVESENKVLWILDSFDPLVAEEDEREVSVPAILRKSRFA 2282  
 QY 623 QALPVWARPDPNPLVETWKKPDYEPVPHG 653  
 Db 2283 PALPVWARPDPNPLVETWKKPDYEPVPHG 2313

RESULT 4  
 S18030  
 genome polyprotein - hepatitis C virus (isolate JKI)  
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
 A;Species: hepatitis C virus  
 A;Variety: isolate JKI  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
 C;Accession: S18030; S33570; A48332; S18029  
 R;Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 A;Submitted to the EMBL Data Library, September 1991  
 A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
 A;Reference number: S18028  
 A;Accession: S18030  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3010 <HON>  
 A;Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479  
 A;Experimental source: isolate JKI from an individual  
 R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
 Arch. Virol. 128, 163-169, 1993  
 A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A;Reference number: A48332; MUID:93119270; PMID:8380322  
 A;Accession: S33570  
 A;Molecule type: genomic RNA  
 A;Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>  
 A;Cross-references: EMBL:X61591  
 A;Note: this sequence is inconsistent with the nucleotide translation  
 A;Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A;Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:P.121748)  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin  
 F;2-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <EPM>  
 F;192-389/Product: major envelope protein E #status predicted <MEE>  
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>  
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F;1312-1317/Region: nucleotide-binding motif A (P-loop)  
 F;1316-1319/Region: DEXH motif  
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
 F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
 F;196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (As

Query Match 34.4%; Score 1534; DB 1; Length 3010;  
 Best Local Similarity 32.3%; Pred. No. 8.3e-87;  
 Matches 397; Conservative 41; Mismatches 115; Indels 676; Gaps 19;  
 QY 146 TDNSSPPVPSQVQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVNLPSVAATLGFAYMSK 205  
 Db 1211 TDNSSPPAVPQTPQVAHLHAPTSGSGKSTKVPAAAYAAQGYKVLVNLPSVAATLGFAYMSK 1270  
 QY 206 AHGIDNIRGTGVTITTTGSPITYTYTGKFLADGCGSGGAYDIIICDCHSTDTATSLIGIG 265  
 Db 1271 AHGVDNISTGVTITTTGAPITYTYTGKFLADGCGSGGAYDIIICDCHSTDTATSLIGIG 1330  
 QY 266 TVLDQAEATAGARLVLATATPPGCVTPVPHNIEEVALSTTGEIPFYGKAIPLEIVIKGRH 325  
 Db 1331 TVLDQAEATAGARLVLAATAATPPGCVTPVPHNIEEVALPNTGEIPFYGKAIPLEIVIKGRH 1390  
 QY 326 LIFCHSKKCKDELAALKVNAAYRGDLVSVIPTSQDVVVVATDALMTGTGDFDS 385  
 Db 1391 LIFCHSKKCKDELAALKVNAAYRGDLVSVIPTSQDVVVVATDALMTGTGDFDS 1450  
 QY 386 VIDCNTC----- 392  
 Db 1451 VIDCNTCVTQTVDFSLDPTFTTITTTLPQDANRSORRGHTGRGCIYRFVTPGERPSG 1510  
 QY 393 ----- 392  
 Db 1511 MPDSSVLCEYDAGCAWYELTPAVTSVRLRAYLNTPLPVCQVHLEFWESVFTGLTHIDA 1570  
 QY 393 ----- 392  
 Db 1571 HFLSQTQAGENPFYLVAYQATVCARAQAPPSPSQDQMKCLIRLKTPLHGTPLLYRLGA 1630  
 QY 393 -----AC-----SG 396  
 Db 1631 VQNEVTLTHPTIFKIMACHSADLEVTVTSTWLVGGVLAALAYCLTGTGTVVIGRIILSG 1690  
 QY 397 KPAAIPDREVLYREFDEMEECSQHLPIYIEQGMMLAEQFKQKALGSLRGKGPAPVDPKEVL 456  
 Db 1691 KPAAIPDREVLYREFDEMEECSQHLPIYIEQGMMLAEQFKQKALGSLRGKGPAPVDPKEVL 1740  
 QY 457 YQYDEMEECSQAAPYIE-----QAQVIAH-----QFKEVGLGLINDQV----- 497  
 Db 1741 -----KQAEAAAPVVESSKQWALEAFWAKHMNFISGIQYLAGLSTLPONPAIVSLMAF 1793  
 QY 498 -----VTPDKILY-----EADDEMEECSKAALIEEGORMAEMLK 533  
 Db 1794 TASITSPLTQHTLFLNLLGGWAAQLAPPSAASAFVGAGIAGAVSGISGLGVLDILA 1853  
 QY 534 SKIQGLLG-----ILRRHVGPGEAV 554  
 Db 1854 GYGAGVAGALVAFKMGSGEMPSTEDLVNLLPALSPALVGVVCAAILRRHVGPGEAV 1913  
 QY 555 QMWNRLIAFASRGNHVSPTHYVP-----SRSRFAQALPWARDYN 596  
 Db 1914 QMWNRLIAFASRGNHVSPTHYVPESDAAARVTKILSSLTITQRLRLHQ-----WINEDCS 1969  
 QY 597 PPLVETWKK----- 605  
 Db 1970 TPCSGSLWRDWDWICITVLTDFKTLQSKLLPRLPGDPFPCORGVRGVRGDMQVMTTC 2029  
 QY 606 -----PDY----- 608  
 Db 2030 PCGAQITGHVKGSMRIVGPKTCSNTHWGTFFPINAVTTGPTSPAPNYSRALWRVAEE 2089  
 QY 609 ----- 608  
 Db 2090 YVEVTRVGDPHYVTGMTDNNVKCPQVPAPEFFTEVDGVLRLHRYAPACKPLLRDEVTFQV 2149  
 QY 609 -----BPPVV----- 613  
 Db 2150 GLNQFPVGSQLPCEPEFDPVTVLTSMLTDPSHITAETAKRLARGSPSSASSASQSLAP 2209  
 QY 614 -----HG----- 615



```

Db      2210 SLKATCTTRDSDPADLIEANLLWRQMGNGNITRVESENKVWILDSFEPLRAEEDEREVS 2269
Qy      616 -----RSSRRFAQALPVMARPDYNPLVETWTKPDYEPVPPVHGRKTKRNTNRRPQDVKF 669
Db      2270 VAAEILKSRKFPFALPIWAPSPNPLLESWKDPDVPVPHGCP-----PPTMAP 2322
Qy      670 PGGQIVGVYLLPRRGR-----LGVLATRKTSPIPKARRPGRRTWAQPGYP 717
Db      2323 P-----IPPRKRKTWLTSTVSSALAEATKTFGSSGSSAVDSGTATAPDQP 2372
Qy      718 WPLYGNKDRSTGKSW-----GKPGYP 739
Db      2373 ---SDDGRGSDDESYSMPPEGEPCDP 2398

RESULT 5
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: s
A;Reference number: A45573; MUID:92295714; PMID:1318627
A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1;
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <BPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepatitis C virus genome polyprotein
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match          34.4%; Score 1532.5; DB 1; Length 3010;
Best Local Similarity 33.5%; Pred. No. 1e-86;
Matches 373; Conserved 38; Mismatches 87; Indels 615; Gaps 13;

Qy      146 TDNSSPPVQSFQVAHLHAPTGSKSTKPAAYAAQYKVLNPSVAATLGFAYMSK 205
Db      1211 TDNSSPPAVPQTFQVAHLHAPTGSKSTKPAAYAAQYKVLNPSVAATLGFAYMSK 1270
Qy      206 AHGIDPNIRTVIRITGTSPTYSTYTGKFLADGCGSGAYDIIICDECHSDTATSIILGIG 265
Db      1271 AHGIDPNIRTVIRITGTSPTYSTYTGKFLADGCGSGAYDIIICDECHSDTATSIILGIG 1330
Qy      266 TVLQDAETAGARLVVLAATATPPGTVVPHNIEEVALSTGTEIFPYGKAILEVYKGRH 325
Db      1331 TVLQDAETAGARLVVLAATATPPGTVVPHNIEEVALSTGTEIFPYGKAILEVYKGRH 1390
Qy      326 LIFCHSKKKCKDELAALKVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGTGDPDS 385
Db      1391 LIFCHSKKKCKDELAALKSLGLGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGTGDPDS 1450
Qy      386 VIDNCNTC----- 392
Db      1451 VIDNCNTQTQTVDFSLDPTFTTITTTVPQDAVRSQRKRTGRGRGGIYRFTVPCRPSG 1510

```

```

Qy      393 ----- 392
Db      1511 MFDSSVLCECYDAGCAGWYELTPAETTVRLRAYLNTPLGLPVCQDHLFEWESVFTGLTHIDA 1570
Qy      393 ----- 392
Db      1571 HFLSQTKAGDNFPYLVAQATVCARAQAPPSPSDQWQKCLIRLKPTLHGPTPLLYRLGA 1630
Qy      393 -----AC-----SG 396
Db      1631 VQNEITLTHPTIKFIMACMSADLEWTVTWLVGGVLAALAAAYCLTTGTVVIVGRIILSG 1690
Qy      397 KPAILIPDREVLYREFDEMESCQHLPIYIEQOMLAEQFKOKALGLSGRGPAPVDPDEV 456
Db      1691 RPAVVPDREVLYREFDEMESCASHLPYIEQOMLAEQFKOKALGLLQATKQAEAAAPVW 1750
Qy      457 YQYDEMEE----- 465
Db      1751 ESRWRALAEAFWAKHMNFISGIQVLAGLSTLPGNPATASLMAPTASITSPLTNTONTLLFN 1810
Qy      466 -----CSQAAPYIEQAQVIAHQFKEKVLGLINDQVVTTPDKEL-----YEAF 509
Db      1811 ILGQWVAQALAPPASAASAFVAGIAGAAIGSIGLVLV---DILAGYGAGVAGALVAF 1866
Qy      510 DEMEECASKA-----ALIEGQRMAMLMKSKIQGLLGLLRHVRHVGEGAVQMMRLI 561
Db      1867 KWSGEAPSAEDLVNLLPAILSPCALVGVV-----CAAILRRHVRHVGEGAVQMMRLI 1920
Qy      562 AFASRGNHVSPTHYVP-----SRSRRFAQALPV-----WAPDYNPLVETWTK-- 605
Db      1921 AFASRGNHVSPTHYVPESDAAARVTQILSSLTITQLKRLHQWINECDSTPCSGSLWKDV 1980
Qy      606 ----- 605
Db      1981 WDMICTVLTDFKTWLOSKLLPKLPGVFFPFCQSGYKGVWRGDGIMQTTCPGQAITGHVK 2040
Qy      606 -----PDY----- 608
Db      2041 NGSMRIVGPKTCSNTWHTGTFPINAITYTTPCTSPAPNYSRALMRVAABEYVEITRVGDFH 2100
Qy      609 ----- 608
Db      2101 YVTGTTNDNVKPCQVQVAPBFFTELDGVLRLHRYAPACRPLLRBVTQVGLNQYLVGSOL 2160
Qy      609 -----EPVW-----HG 615
Db      2161 PCEPEPDVAVLTSLMTDPSHITAKRLARGSPSSLASSASQLSAPSLKATCTTHD 2220
Qy      616 -----RSRR 620
Db      2221 SPDADLIEANLLWRQMGNGNITRVESENKVWILDSFDPLRAEEDEREVSAAEILRKSKK 2280
Qy      621 FAQALPVMARPDYNPLVETWTKPDYEPVPPVH 653
Db      2281 PPPALPIWAPPDYNPLLESWKSPDYVPPVAVHG 2313

RESULT 6
GNWVC
genome polyprotein - hepatitis C virus
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
C;Species: hepatitis C virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004
C;Accession: A38465
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A;Title: Structure and organization of the hepatitis C virus genome isolated from human
A;Reference number: A38465; MUID:91140698; PMID:1847440
A;Accession: A38465
A;Molecule type: genomic RNA
A;Residues: 1-3010 <TAK>

```

A:Cross-references: UNIPROT:P26663; EMBL:M59335; NID:g329770; PID:AAA7945.1; PID:g329770

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein

F:2-115/Product: capsid protein C #status predicted <CPC>

F:116-191/Product: envelope protein M #status predicted <BPM>

F:192-389/Product: major envelope protein E #status predicted <MEE>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1207-1615/Product: hepatitis virus #status predicted <NS3>

F:1230-1237/Region: nucleotide-binding motif A (p-loop)

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2222

	Query Match	34.2%	Score 1523.5	DB 1	Length 3010
	Best Local Similarity	33.2%	Pred. No. 3.8e-86		
	Matches 372	Conservative	39	Mismatches 80	Gaps 14
Qy	146	TDNSSPPVVFQSFQVAHLHAPTSGSGKSTKVPAAQAQGYKVLNPNPSVAATLFGAYMSK	205		
Db	1211	TDNSSPPVVFQSFQVAHLHAPTSGSGKSTKVPAAQAQGYKVLNPNPSVAATLFGAYMSK	1270		
Qy	206	AHGIDPNIRTVRTITITGSPITVITYCKKFLADGCGSGGAYDIIICDECHSDTATSILGIG	265		
Db	1271	AHGIDPNIRTVRTITITGSPITVITYCKKFLADGCGSGGAYDIIICDECHSDTATSILGIG	1330		
Qy	266	TVLDAQETAGARLVLATATPPGSGVTPHPNIEEVALSTTGEIPFYKKAIPLEVIKGGRH	325		
Db	1331	TVLDAQETAGARLVLATATPPGSGVTPHPNIEEVALSTTGEIPFYKKAIPLEAIRGGRH	1390		
Qy	326	LIFCHSKKCDLAALKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGYTGDFDS	385		
Db	1391	LIFCHSKKCDLAALKLGLGINAVAYRGLDVSIVPTIGDVVVVATDALTMTGYTGDFDS	1450		
Qy	386	VIDCMTCTCTVDFSLDPTFTIETTVQDAVRSQRGRGRTYRFTVTPGERPSG	1510		
Db	1451	VIDCMTCTCTVDFSLDPTFTIETTVQDAVRSQRGRGRTYRFTVTPGERPSG	1510		
Qy	393	-----AC-----	392		
Db	1511	MPDSSVLCYCYDAGCANYELTPTAETSVRLRAYLNTPGLPVCQDHLFEWESVPTGLTHIDA	1570		
Qy	393	-----	392		
Db	1571	HFLSOTKQAGDNFPYLVAYQATVCARAQAPPPQDQMKLIRLKTLLHGPTFLLYRLQA	1630		
Qy	393	-----AC-----	396		
Db	1631	VQNEVTLTHPIITYKIIMACMSADLEVTSTVVLGVLAALAAAYCLTTSVVIVGRILLSG	1690		
Qy	397	KPAIIPDREVLRYREFDEMEECSOHLPTIEOGMLAEQFKQKALGSRGGKPAIVDPKEVL	456		
Db	1691	REPAIVPDRELLYEFDEMEECSAHLPTIEOGMLAEQFKQKALGELQ-----	1737		
Qy	457	YQYDEMEECSQAAPYIEQAQVIAHQFKEK-VLGLINDQVV-----	497		
Db	1738	-----TATKQAAAPVYESKWRALETTFWAKHMWNFISGIQYLAGLSTLPGNPAIASLMAF	1793		
Qy	498	-----VTPDKELIY-----	533		
Db	1794	TASITSPLTTQSTILLFNILGGVAAQLAPPSSAASAFVGCAGIAGAAGVSGTGLKVLVDILA	1853		
Qy	534	SKIQGLLG-----	554		
Db	1854	GYGAGVAGALVAFKWSGEMPTSEDVNLNLPAILSPGALVGVVCAAILRRHVGPGEVAV	1913		
Qy	555	QNMNRLIAFASRGNHVSPTHYVPS-----	578		
Db	1914	QNMNRLIAFASRGNHVSPTHYVPSDAARVTOILSSITITQLLKRHLQHWINEDCSTPCS	1973		

```
QY 212 NIRTGVRTTTGSPITVSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDQA 271
DB 1277 NIRTGVRTTTGAPITVSTYKFLADGCGSGAYDIIIMCDECHSTDDSTTILGIGTVLDQA 1336
QY 272 ETAGARLVVLTATPGSVTVPHNIEEVALSTTGEIPFYKAIPLFVKGGRHLIFCHS 331
DB 1337 ETAGARLVVLTATPPGSVTVPHNIEEVALSTTGEIPFYKAIPIETIKGGRHLIFCHS 1396
QY 332 KKKCDELAALKLVALGINAVAYRGLDVSVIPTS GDVVVVATDALMTGVTGDFSDVIDCNT 391
DB 1397 KKKCDELAALKSALGINAVAYRGLDVSVIPASGNVVVATDALMTGVTGDFSDVIDCNT 1456
QY 392 C----- 392
DB 1457 CVTQTVPFSLDPTFTIETTTMPQDAVSRSORRGRTSRGRGIYRFVTPGERSPGMFDSSV 1516
QY 393 ----- 392
DB 1517 LCECYDAGCAWYELTPAETSRLRAYLNTPLGVPQCQDHLFWESVFTGLTHIDAHFLSQT 1576
QY 393 ----- 392
DB 1577 KOAGDNFPYLVAYQATVCARAQAPPSPWDQWKKCLTRLKPTLHGPTPLLYRLGAVQNEVT 1636
QY 393 -----AC-----SGKPAIIP 402
DB 1637 LTHPTIKYIMACSADLEVTVSTWLVGVGLAALAAAYCLTTGVSIVVIGRIILSGKPAVVP 1696
QY 403 DREVLRYRFEDEMECSQHLPIYEQGMALAEQFKQKALGLSGRGKPAIVDPDKVLYQQYDE 462
DB 1697 DREVLRYRFEDEMECSQHLPIYEQGMALAEQFKQKALGLSQ-----TA 1739
QY 463 MEECSQAAPYIEQAQ-----VIAHQPKKVLGLINDQVV----- 497
DB 1740 TKQAEAAAPVESKWRTELEAFWANDMWNFISGIVYLAGLSTLPGNPATSLMAFTASITS 1799
QY 498 -VTPDKELLY-----EAFDEMECSKAALIERGORMAEMLSKIQGL 539
DB 1800 PLTTQSTLLFNILGGVAAQLAPPQGAASAFVGCAGIAGAAGVSGIGLKVLDVMVAGYGAV 1859
QY 540 LG-----ILRRHVGPGEGAVQMMNRL 560
DB 1860 AGALVAFKVMGSEMPSTEDLVNLLPAILSPGALVGVVCAAILRRHVDPGEGAVQMMNRL 1919
QY 561 IAFASRGNHVSPTHVPS-----RSRRFAQA----- 578
DB 1920 IAFASRGNHVSPTHVPSDAAARVTQILSLTITOLLRLHQNEDCSTPCSGSLRD 1979
QY 579 ----- 578
DB 1980 VWDWICTVLADFKTWLQSKLLPRLPGVPFFSCQGYKGVWRGDGIMQITCPCGAQLTGHV 2039
QY 579 ----- 578
DB 2040 KNGSMRIWGPKTCSTNTHGTPPINAYTTGCTPSPAPNYSRALWRVAEEYVEVRRVGDF 2099
QY 579 -----RSRRFAQA----- 586
DB 2100 HYVTGMTDNVCKPCQVPAPEFFTEVDGVRHLRYAPACKPLLRREEVSFQVGLNQYVVGSG 2159
QY 587 LPVWARPD-----Y 595
DB 2160 LPCEPEPDVAVLTSMLTDPSSHITAETAKRRLARGSPPSLASSASQLSALSKAACTTRH 2219
QY 596 NPP---LVET---WKK-----PDYEP-----PVVHGRSSR 619
DB 2220 TTPDADLIEANLLRQEMGNITRVESENKVILDSFDFLRAEEDEREVSVPFAEILRKSR 2279
QY 620 RFAQALPVWARPDYNPPLVETWKKPDYPPVHVG 653
DB 2280 KPFPALPVWARPDYNPPLLEPKWDPDYPPVHVG 2313
```

## RESULT 8

GNWVCJ

Genome polyprotein - hepatitis C virus (strain J)

N;Contains: capsid protein C; envelope protein E; major envelope protein M; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C;Accession: A39253; PS0086

R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto

Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990

A;Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients

A;Reference number: A39253; MUID:91088550; PMID:2175903

A;Accession: A39253

A;Molecule type: genomic RNA

A;Residues: 1-3010 &lt;KAT&gt;

A;Cross-references: UNIPROT:P26662; GB:D90208; NID:G221610; PID:BAA14233.1; PID:G221611

R;Kato, N.; Ohkoshi, S.; Shimotohno, K.

Proc. Jpn. Acad. 55B, 219-223, 1989

A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A;Reference number: PS0085

A;Accession: PS0086

A;Molecule type: genomic RNA

A;Residues: 2650-2707 &lt;KA2&gt;

A;Experimental source: Japanese isolate

C;Comment: The cleavage sites of this polyprotein have not been determined.

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin

F;2-115/Product: capsid protein C #status predicted &lt;CPC&gt;

F;116-191/Product: envelope protein M #status predicted &lt;EPM&gt;

F;192-389/Product: major envelope protein E #status predicted &lt;MSE&gt;

F;390-729/Product: nonstructural protein NS1 #status predicted &lt;NS1&gt;

F;730-1006/Product: nonstructural protein NS2 #status predicted &lt;NS2&gt;

F;1007-1615/Product: hepatitis virus protein NS3 #status predicted &lt;NS3&gt;

F;1230-1237/Region: nucleotide-binding motif A (P-loop)

F;1312-1317/Region: nucleotide-binding motif B

F;1316-1319/Region: DEXH motif

F;1616-1862/Product: nonstructural protein NS4a #status predicted &lt;NS4a&gt;

F;1863-2013/Product: nonstructural protein NS4b #status predicted &lt;NS4b&gt;

F;2014-3010/Product: nonstructural protein NS5 #status predicted &lt;NS5&gt;

F;196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2

Query Match 34.1%; Score 1519; DB 1; Length 3010;

Best Local Similarity 32.0%; Pred. No. 7.2e-86;

Matches 386; Conservative 44; Mismatches 108; Indels 670; Gaps 18;

QY 115 HVTG-GAARTSGTSL--FSPGASQNIQLITS--TDNSPPVVPVQSFQVAHLHAPTGS 169

DB 1175 HVGIFRAAVCTRGVAKAVDFIPVESMETTWRSVPVTDNSPPVAVPQTFQVAHLHAPTGS 1234

QY 170 GKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAGYMSKAHGIDFNIRTGVRTITGSPITYS 229

DB 1235 GKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAGYMSKAHGIEPNIRTGVRTITGSPITYS 1294

QY 230 TYGFLADGGCGGAYDIIICDECHSTDATSIILGIGTVLDQAGARLVLTATPPGS 289

DB 1295 TYCKFLADGGCGGAYDIIICDECHSTDTSTLIGTIGTVLDQAGARLVLTATPPGS 1354

QY 290 VTVPHNTEEVALSTTGEIPFYKAIPLFVKGGRHLIFCHSKKCKDELAALKLVALGINA 349

DB 1355 ITVPHNTEEVALSTTGEIPFYKAIPIEATKGGRRHLIFCHSKKCKDELAALKLVGLNA 1414

QY 350 VAYYRGLDVSVIPTS GDVVVVATDALMTGVTGDFSDVIDCNTC----- 392

DB 1415 VAYYRGLDVSVIPTS GDVVVVATDALMTGVTGDFSDVIDCNTCTVQTVDPSLDPTFTIET 1474

QY 393 ----- 392

DB 1475 TTLPDQAVSRAQRGRGRTGRSGIYRFVTPGERSPGMFDSSVLCBYDAGCAWYELTPAE 1534

QY 393 ----- 392

DB 1535 TSVRLRAYLNTPLGVPQCQDHLFWESVFTGLTHIDAHFLSQTKQAGDNLPYLVAYQATVC 1594

QY	393	-----AC-----	394
Db	1595	ARAQAPPSMDQMKCLRLKPTLHGPTPLLYLGGAVQNEVTLTHPTKYIMACMSADLE	1654
QY	395	-----SGKPAIPDPREVLYREFDEMECSQH	420
Db	1655	VVTSTWLVGVLAAALAAAYCLTTGTSVVIVGRPIILISGRPAVIPDPREVLYQEFDEMEECASH	1714
QY	421	LPVIEQGMMLAEQPKQALGLSRGKPAIPDPKEVLYQQYDEMEECSSQAAPYIEQAQ---	477
Db	1715	LPVIEQGMMLAEQPKQALGLLQ-----TATKQAEAAAAPVVBESKWRAL	1757
QY	478	-----VIAHQPKFKEVLGLINDQVW-----VTPDKELIY-----	506
Db	1758	EVFWAKHMMNFISGIIQYLAGLSTLPGNPAIASLMAFTASITSLTNTQTLFNLIGGWA	1817
QY	507	-----EAPDEMECSKAALIERGORMAEMLSKIOGLIG-----	541
Db	1818	AQLAPPSAASAFVAGIAGAAVSGIGLGKVLVDILAGYGAGVAGALVAFKVMGSEMPSTE	1877
QY	542	-----ILRRHVGPGEGAVQMMRLIAFASRGNHVSPTHVP-	577
Db	1878	DLVNLPAIILSPGALVGVVCAAILRRHVGPGEGAVQMMRLIAFASRGNHVSPTHVPE	1937
QY	578	-----SRSRFPAQALPV-----WAPDPYNPPLVETWKK-----	605
Db	1938	SDAARVTQILSSITITQLLKRHLQWINEDCSTPCSGSLWKVDWIDICTVLSDPKTWLQS	1997
QY	606	-----PDY-----	605
Db	1998	KLLPRLPGLPFLSCQGYKGVMRGDGMQTTCPGQAQITGHVKNMSMRIVGPKTCNTWH	2057
QY	606	-----PDY-----	608
Db	2058	GTPPINAYTTCPTSPAPNYSRALWRVAABEYVEVTRVGDHFHYVTGTTDNDNVKPCQVP	2117
QY	609	-----EPPVV-----	613
Db	2118	APFEFTEVDGVRLLHRYAPVCKPLLRBEVWFQVGLNQYLVGSQLPCEPEPDVAVLTSMLTD	2177
QY	614	-----HG-----	615
Db	2178	PSHITAETAKRRLARGSPPLSSASSASQLSAPLSUKATCTTHHDSFPADLIEANLLWRQEM	2237
QY	616	-----RSSRFAQALPVWARPDPYNPPL	637
Db	2238	GGNITRVESENKVVILDSFPIRAVEDERELSVPAEILRPRKPPPALPIWARPDPYNPPL	2297
QY	638	VETWKKPDYBPPVHVGRKTKRNTNRRPQDKFPGGQIVGGVYLLPRRGRPLGLVLRKT	697
Db	2298	LESWKDPDYVPPVHG-----CP-----LPSTKA	2321
QY	698	SPIPKARR	705
Db	2322	PIPPPPRR	2329

RESULT 9

JQ1303

genome polyprotein - hepatitis C virus (isolate HC-J6)

N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004

C;Accession: JQ1303

R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Miyakawa, Y. J. Gen. Virol. 72, 2697-2704, 1991

A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human

A;Reference number: JQ1303; MUID:92044440; PMID:1658196

A;Accession: JQ1303

A;Molecule type: genomic RNA

A;Residues: 1-3033 <OKA>

A;Cross-references: UNIPROT:P26660; GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651			
A;Experimental source: isolate HC-J6 from a Japanese individual			
C;Superfamily: hepatitis C virus genome polyprotein			
C;Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmembrane			
F;2-115/Product: capsid protein C #status predicted <CPC>			
F;116-191/Product: envelope protein M #status predicted <EPM>			
F;192-389/Product: major envelope protein E #status predicted <MES>			
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>			
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>			
F;1011-1619/Product: hepatitis C virus nonstructural protein NS2 #status predicted <NS2>			
F;1316-1321/Region: nucleotide-binding motif B			
F;1320-1323/Region: DEXH motif			
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>			
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>			
F;2018-3033/Product: nonstructural protein NS5 #status predicted <N05>			
F;196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,281			
Query Match 32.1%; Score 1429.5; DB 1; Length 3033;			
Best Local Similarity 31.2%; Pred. No. 2.7e-80;			
Matches 393; Conservative 62; Mismatches 165; Indels 639; Gaps 27;			
QY	26	SNGPV-KVMGSIKGLTEGLHGFHFHFEFGDNTAGCTSAGPHFNPLSTRGCNCSTVPGHITG	84
Db	1097	SRGPVTQMYSASG--DLVGPSPGCTKSLEPCTCGAVDLY-LVTR--NADVIPARRRG	1150
QY	85	HRMAWKLG----SAAATTSGFVSLFAPGAKQNEHTVGTGAAARTTSGLTSLSPGASQNI	140
Db	1151	DKRGALLSPRLSTLKGSSGGVPLCPRG-----HAGGVFRA-----AVCSRGVAKSEI	1197
QY	141	QLI-----TSTNSPPVVPVPOSFOVAHLHAPTSQSGKSTKVPAAAYAAQGYKVLV	188
Db	1198	DFIPVETLDIVTRSPFSDNSTPPAVPQTVQVGLHAPTSQSGKSTKVPVAYAAQGYKVLV	1257
QY	189	LNPSVAATLFGAYMSKAHGINRTGVRTITTTGSPITVSTYTKFLADGCGSGGAYDII	248
Db	1258	LNPSVAATLFGAYLSKAHGINRTGVRTITTTGSPITVSTYTKFLADGCGGAGAYDII	1317
QY	249	ICDECHSTDATSLIGITGTVLDQATAGARLVLAATATPPGSVTVPHNIEVALSTTGEI	308
Db	1318	ICDECHAVDSTTILGITGTVLDQATAGVRLTVLAATATPPGSVTTPHPNIEVALGQGEI	1377
QY	309	PFTGKAIPLEVIKGGHILFCHSKKCDLAALVALGINAVAYYRGLDVSIVPTSGDVV	368
Db	1378	PFTGKAIPLEVIKGGHILFCHSKKCDLAALRGMLNAVAYYRGLDVSIVPTQGDVV	1437
QY	369	VVATDALMTGTCGDFSDVDCNT-----	391
Db	1438	VVATDALMTGTCGDFSDVDCNVAVTVQVDFSLDPTFTTTTQVTPQDAVSRQRGTGR	1497
QY	392	-----CAC--SG-----KPAIIPDREVLYREFDEMEECSQ	419
Db	1498	GRLGIYRYVSTGERASGMFDSVVLCEYDAGAAWYELTPAETTVRLRAYFNTPGVPVQCD	1557
QY	420	HLP-----YIEQGMMLAEQF-----KQALGLSRGG-----	445
Db	1558	HLFEFWEAVFTGLTHIDAHFLSQTQSGENFAYLTAYQATVCARAKAPPPSDVMWKCLTR	1617
QY	446	-KPAIVDPKEVLYQQYDEMEECSQAAP---YIEQ-----AQVIAH	481
Db	1618	LKPTLVGPTPLLYRLGSVTNEVTLTHPTVKYIATCWAQDLEVMTSTWVLAGGVLAAVAAY	1677
QY	482	QFKEKVLGLID---NDQVVTTPDKELIYEAFAFDEMEECSKAALIEEGQMAEMLKSKIQ	537
Db	1678	CLATGCVCIIGRLHVNQRAVAVPDKVLYEAFAFDEMEECSRAALIEEGQMAEMLKSKIQ	1737
QY	538	GLL-----	540
Db	1738	GLLQASKQAQDIQPAVQASWPKEQFWAKHMMNFISGIQYLAGLSTLPGNPAVAMMAF	1797
QY	541	-----	540
Db	1798	SAALTSPSTSTTILNIGGLWLASQIAPPAGATGTFVWSGLVGAANVGSIGLGVLDILA	1857

```
Qy 541 -----GILRRHVGPGEAV 554
Db 1858 GYGAGISGALVAFKIMSGKXPMEDVVNLLPGILSPGALVVGVCIAILRRHVGPGEAV 1917
Qy 555 QMWNRLIAFASRGNHVSPTHYV--PSRSRFAQAL----- 587
Db 1918 QMWNRLIAFASRGNHVAPTHYVTESDASQRTQLGSLTITSLRLRHWNWITEDCIPCS 1977
Qy 588 -----PVNA----- 591
Db 1978 GSWLRDWDVVCITLTDKKNLTKLPFKMPLPFISCKGKGVAGTGIMTTRCPGA 2037
Qy 592 -----RDYNPLVETWK----- 604
Db 2038 NISGNVRLGSMRITGPKTCMNIWOGTPIINCYTEGQCVPKPAFNK---IAIWRVAASEY 2094
Qy 605 -----KP----- 606
Db 2095 AEVTQHSYHYITGLTDDNLKVPCLPSPFFSWDVGQIHRFAPIPKPFPRDEVSPCVG 2154
Qy 607 -----DYPE----- 610
Db 2155 LNSFVGSQLPCEPEPTDVLTSMLTDPSSHITAARLARGSPPEASSASQLSAPS 2214
Qy 611 -----PVVHGRS----- 617
Db 2215 LRATCTTHGKAYDVMVDANLFMGGDVTRIESESKVVLDSLDPWVEERDLEPSIPSEY 2274
Qy 618 ---SRRAQALPWARDYNPLVETWKDYPVPHG-----RKTENTNRRPODV 667
Db 2275 MLPKREPPALPAWAPDYNNPLVESWKRPDYQATVAGCALPPKKTPTPPRRRRRTV 2333

RESULT 10
GNWVJ8
genome polyprotein - hepatitis C virus (strain HC-J8)
N;Contents: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J8)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: A40250; PQ0397; F00559
R;Okamoto, H.; Kuri, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to repc
A;Reference number: A40250; MUID:92230232; PMID:1314459
A;Accession: A40250
A;Molecule type: Genomic RNA
A;Residues: 1-3033 <OKA>
A;Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:G221608; PIDN:BAA01761.1;
R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I.
J. Gen. Virol. 73, 1131-1141, 1992
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A;Reference number: PQ0393; MUID:92268871; PMID:1316939
A;Accession: PQ0397
A;Molecule type: Genomic RNA
A;Residues: 2678-2754 <CHA>
A;Cross-references: DDBJ:D10134
A;Experimental sources: Isolate E-b12
R;Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A;Title: Distribution of plural HCV types in Japan.
A;Reference number: PQ0554; MUID:92068204; PMID:1720309
A;Accession: PQ0559
A;Molecule type: mRNA
A;Residues: 2678-2729 <KAT>
A;Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BAA01418.1; PID:G221524
C;Superfamily: hepatitis C virus genome polyprotein
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F;1-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
```

```
F;1011-1619/Product: hepatitis C virus #status predicted <NS3>
F;1234-1241/Region: nucleotide-binding motif A (P-loop)
F;1316-1321/Region: nucleotide-binding motif B
F;1320-1323/Region: DEXH motif
F;1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F;1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F;2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F;196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,231
```

```
Query Match 31.9%; Score 1422.5; DB 1; Length 3033;
Best Local Similarity 29.6%; Pred. No. 7.4e-80;
Matches 408; Conservative 62; Mismatches 182; Indels 725; Gaps 31;
```

```
Qy 22 EQKESNGPVKWSIKGLTEGLHGHV-----HEFGDNTAGCTTSAGHPPLSTRGC 73
Db 1055 DKNEAQGVQLSSV---TQFLGTSISGLVMTVYHAGNKTLL---AGP-----KGP 1100
Qy 74 NCSIIYFCHITGHRMANKLGSAAAT---TSGFYSLF-----AP 107
Db 1101 VTQMYTS-AGDVLVGWPPSPPTKSLDPCTCGADVLLVLTNRNADVIPVRRKDDRRGALLSP 1159
Qy 108 GAKQNETHTVTGGAARTTSLTSLF-----SPGASONIQLI-----TSTDNS 150
Db 1160 RPLSTLKGSSGGPVLCSRGHAGVGLFRAAVCAVCAKSIDFIPVESLDVATRTSPFSDNST 1219
Qy 151 PPVVPQSFOVAHLHAPTSGSKSTKVPAAAYAAQGVKVLVLPNSVAATLFGAYMSKAHGD 210
Db 1220 PPAPVQSYQVGYLHAPTSGSKSTKVPAAAYAAQGVKVLVLPNSVAATLFGAYMSKAHGIN 1279
Qy 211 PNRTGVRTTITGSPITYSTYTGKFLADGGCGGAYDIIICDECHSTDATSLIGTGTVLDDQ 270
Db 1280 PNRTGVRTTITGSDITYSTYTGKFLADGGCGGAYDIIICDECHSDVATTILGIGTGTVLDDQ 1339
Qy 271 AETAGARLVVLATATPPGVSIVTPHPNIEEVALSTTGEIPFYKAIPLKVIKGRHLIFCH 330
Db 1340 AETAGVRLVVLATATPGTPTVTPHSNIEEVALGHEGEIPFYKAIPLAFIKGRHLIFCH 1399
Qy 331 SKKKDEBLAAKLVALGINAVAYRGLDVSIVTSGDVVVVATDALTMTGYGDFDSVDCN 390
Db 1400 SKKKDEBLAAALRGMGVNAVAYRGLDVSIVTQGVVVVATDALTMTGYGDFDSVDCN 1459
Qy 391 TC-----ACSG-KPAIIIPDRE 405
Db 1460 VAVSQIVDFSLDPTFTTTTQVPODAVRSQRGRTGRLGVYRVYSSSERPSGCMFDSV 1519
Qy 406 VLYREFDE-----MEECQHLPIYIE---QGMN-----LAEQ 433
Db 1520 VLCECYDAGAAWYELTPAETTVRLRAYFNTPGLPVCODHLEFWEAVFTGLTHIDAHFLSQ 1579
Qy 434 FKQKALGLSRGG-----KPAIVDPKEVLYQQ 459
Db 1580 TKQ-----GGENFAYLTAQATVCARAKAPPPSWDMWKCLTRLKPTLTGTPTLLYRL 1632
Qy 460 YDEMECSQAAP---YIEQ-----AOVIAHQFKEKVLGLID--- 492
Db 1633 GAVTNEVTLTHPTVKYIATCQADLEIMTSSWSVLGGVLAATAVATCISIIIGRLHL 1692
Qy 493 NDOVVTPDKIELIYEAPEDEMEECASKAALIEEGORMAEMLSKTIQGLL----- 540
Db 1693 NDRVVVAPDKIELIYEAPEDEMEECASKAALIEEGORMAEMLSKTIQGLLQOATROAQDIQP 1752
Qy 541 ----- 540
Db 1753 AIQSSWPKLEQFWAKHWNFTSGIOYLAGLSTLPGNPAVASMMAFSAALTSPLPTSTTL 1812
Qy 541 ----- 540
Db 1813 LNIIMGWLASQIAPPAGATGFWVSGLVAAVSGTGLGKILVDVLVAGYAGISGALVAPKI 1872
Qy 541 -----GILRRHVGPGEAVQVMNRLIAFASRGNH 569
Db 1873 MSGKPTVEDVNNLLPAILSPGALVVGVCIAILRRHVGPGEAVQVMNRLIAFASRGNH 1932
```

QY	570	VSTHYV--PSRRRAQALP	-----	588
Db	1933	VAPTHYVESDASQVTVQLSSLTITSLRLHLAWITEDCPVPCSGSLQDIWDWYCSIL	1992	
		-----VWA-----		
QY	589	-----VWA-----	591	
Db	1993	TDFKNWSSKLLPKMPGIPPISCQGYGVAAGTGWVTRPCGANISGHRMGTWKITG	2052	
		-----RPDYNPLVET--WK-----		
QY	592	-----RPDYNPLVET--WK-----	604	
Db	2053	PKTCLNLWQGTTPINCYTEGPCVKP--PPNYKTAIRVAAASEYVEVTHGHSFSYVTGL	2109	
		-----K-----		
QY	605	-----K-----	605	
Db	2110	TSDNLVKPCQVPAPEFFFSWVDGVQIHRFAPVPGPFRRDEVFTFTVGLNSFVGSQLPDCE	2169	
		-----PPVHGSS-----		
QY	606	PDVE	618	
Db	2170	PDEVLASMLTDPSSHITAEMAAARLRAGSPSQSSASQSLSPSLKATCTTHKTYDCD	2229	
		-----RRFAQALPVWA		
QY	619	MVDANLFMGDVTRESDSKVIVLSDLSMTVEDDREPSVPSEYILKRRKFPALPPWA	2289	
Db	2230	RPDYNPLVETWKPYEPPVHVHGRKTKRNRRPDVFPGGQIVGGVYLLPRRGPRL	689	
		-----		
QY	630	RPDYNPLVETWKPYEPPVHVHGRKTKRNRRPDVFPGGQIVGGVYLLPRRGPRL	689	
Db	2290	RPDYNPLVETWKPYEPPVHVHGRKTKRNRRPDVFPGGQIVGGVYLLPRRGPRL	2313	
		-----PLYGNKDR-RSTGKSMG		
QY	690	GVLATRKTSIPKARPEGHTWAQPGYPW-----	734	
Db	2314	CALPPTPTQTPVPPVPPRRRAKVLTDQNVGVLRMAKVLSPLODNDNSGHSTGADTG	2370	

RESULT 11

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)

N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004

C;Accession: JC5620

A;Reference number: JC5620; MUID:97366593; PMID:9223423

A;Molecule type: mRNA

A;Residues: 1-3014 <CHA>

A;Cross-references: UNIPROT:O39928; GB:Y13184

A;Experimental source: genotype 5a, which predominates in South Africa

A;Note: the translation of the nucleotide sequence is not complete in this paper

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F;2-115/Product: capsid protein C #status predicted <CPC>

F;116-191/Product: envelope protein M #status predicted <EPM>

F;192-389/Product: major envelope protein E #status predicted <MEE>

F;384-408/Region: hypervariable #status predicted

F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>

F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>

F;1008-1616/Product: hepatitis C virus #status predicted <NS3>

F;1231-1238/Region: nucleotide-binding motif A (P-loop)

F;1313-1318/Region: nucleotide-binding motif B

F;1317-1320/Region: DEXH motif

F;1617-1863/Product: nonstructural protein NS4a #status predicted <N4a>

F;1864-2014/Product: nonstructural protein NS4b #status predicted <N4b>

F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>

F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 30.0%; Score 1337; DB 1; Length 3014;

Best Local Similarity 28.9%; Pred. No. 1.5e-74;

Matches 359; Conservative 69; Mismatches 155; Indels 660; Gaps 21;

QY 92 GSAARTTSGVSLF-----APGAKQNET-----HVT 117

Db 2199 LANSASQSLAPSLKATCTQGHHPDADLIKANLLWRCMGNGNTRVEAKNKVILDCFK 2258  
Qy 621 -----FAQLPVPWADYNNPLVETWKKPDYEPVHVGRKTK 657  
Db 2259 PLKEEDDREISVADCFKGPAPFPALPVWAPGYDPPLETTWKRPDYDPPQWGPC-- 2316  
Qy 658 RNTNRRPQDVKFGGGQIVGVYLLPRGPRGLVLAIRKTSPI 700  
Db 2317 -----IPGAGPPVPPLPRKRKXPM 2335

RESULT 12  
S68016  
ATPase/RNA helicase - hepatitis C virus (fragment)  
C:Species: hepatitis C virus  
A:Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: S68016  
R:Jin, L.; Peterson, D.L.  
Arch. Biochem. Biophys. 323, 47-53, 1995  
A:Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
A:Reference number: S68016; MUID:96019946; PMID:7487072  
A:Accession: S68016  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-386 <JIN>  
A:Cross-references: UNIPROT:Q04045  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polyprotein  
F:24-31/Region: nucleotide-binding motif A (P-loop)  
F:86-91/Region: nucleotide-binding motif B  
F:90-93/Region: DEXH motif

Query Match 25.3%; Score 1128.5; DB 2; Length 386;  
Best Local Similarity 89.1%; Pred. No. 7.9e-63;  
Matches 221; Conservative 4; Mismatches 2; Indels 21; Gaps 2;  
Qy 146 TDN-SSPPVQSFQVAHLHAPTSGKSTKVPAAAGQYKVLVNLNPSVAATLFGAYMS 204  
Db 4 TDNSSPPVQSFQVAHLHAPTSGKSTKVPAAAGQYKVLVNLNPSVAATLFGAYMS 63  
Qy 205 KAHGIDPNIRTCVRIITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGI 264  
Db 64 KAHGVD-----YKFLADGCGSGAYDIIICDCHSTDATSIIGI 103  
Qy 265 GTVLDAQETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVIKGR 324  
Db 104 GTVLDAQETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVIKGR 163  
Qy 325 HLI FCHSKKCKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGTGDFD 384  
Db 164 HLI FCHSKKCKDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALTMTGFGDFD 223  
Qy 385 SVIDCNTC 392  
Db 224 SVIDCNTC 231

RESULT 13  
PC2219  
polypeptide - hepatitis C virus (type 5a) (fragments)  
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote  
C:Species: hepatitis C virus  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: PC2219  
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the H  
A:Reference number: PC2219; MUID:94338342; PMID:7520237  
A:Accession: PC2219  
A:Molecule type: mRNA  
A:Residues: 1-876 <STU>  
A:Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579  
A:Experimental source: serum

C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: Glycoprotein  
F:1-191/Product: core #status predicted <COE>  
F:68-78/Region: variable  
F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F:248-338/Region: E2  
F:339-411/Region: NS1 (amino end)  
F:412-783/Product: NS3 #status predicted <NSR>  
F:784-837/Product: NS4 #status predicted <NSA>  
F:838-876/Product: NS4B #status predicted <NSB>  
F:281,287,294,312,340/Binding site: carbohydrate (Aan) (covalent) #status predicted

Query Match 21.4%; Score 954.5; DB 2; Length 876;  
Best Local Similarity 34.6%; Pred. No. 1.7e-51;  
Matches 252; Conservative 57; Mismatches 146; Indels 273; Gaps 18;  
Qy 32 VWGSIKGLTEGL-HGFHVHEFGDNTAGCTAGPHFNPLSTRGCNCSIVPGH----- 81  
Db 140 VGGPIGVARALANGVRVLEGVNATGNLPGCSFSFILALLSCLTVPASAPTTALLVA 199  
Qy 82 -----ITGHRMAWKL-----GSAARTTSGFVSLFAPGAKQNETHVTGGAAA 122  
Db 200 QLLRIPQVVIDIIAGSH--WGVFAAAYASVANWTKVVLVFLFAGVDATTQISGSSA 257  
Qy 123 RTTSGLTSLSPGASQNIQLITSD----NSPPVQSFQVA-----HLHAPTSGSKS 172  
Db 258 QTTTGIASTFTRGAQQLQTLINTNGSMHINRTALNCNDSLOTGTGFIAGLYFYHKFNSSGCP 317  
Qy 173 TKVPA---AYAAQGYKVL---VLNPS-----VAA 195  
Db 318 DRMASCRALAFDDQGWGTISYANISGFSDDKPYCHVHPPRPGVVPVPAQVCGVYCFTPS 377  
Qy 196 TLGFGAYMSKAHGDIPNIRTCVRT-----ITGSPITYSTYCKFLADGCGSGG 243  
Db 378 PVVVGTTDSKGH---PTVNWGSNVTDFPLMNTNRPITGASITYSTYCKFLADGCGSGG 434  
Qy 244 AYDIIICDCHSTDATSIIGITGVLDAQETAGARLVVLTATPPGVTVPHPNIEVALS 303  
Db 435 AYDIIICDCHSQDATTILGIGTGLDQAEAGARLVVLTATPPGVTVPHPNIEVALP 494  
Qy 304 TTGEIPFYGKAIPLVIKGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVSIVPT 363  
Db 495 QEGEVFPFYGRAIPLAFIKGGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVAIPT 554  
Qy 364 SGDVVVVATDALTMTGTGDFSDVDCNTC----- 392  
Db 555 AGDVVVCSTDALMTGTGDFSDVDCNSAVTQTVDFSLDPTTIBTTTVPQDAVRSQRR 614  
Qy 393 -----ACSGKPAIIPDREVLVYREFDE-----M 414  
Db 615 GRTGRGRHGIYRVVSAGERPSDMFDSVVLCEYDAGCAWYDLTPAETTVRLRAYINTPGL 674  
Qy 415 EECQHLPLYE-----QG-----WMLAQ-----FKOKA----- 438  
Db 675 PVCQDHLFEWEGVFTGLTNIDAHMLSQTKQGENPPYLVAIYQATVCVRAKAPPSPSHDTMW 734  
Qy 439 -----LG----- 440  
Db 735 KCMLEKLTLTGPTPLLYRLGPVQNEITLTHPIKTYIMACMSADLEVITSTWLVGVVVA 794  
Qy 441 -----LSRGGKPAIVDPKKEVLVYQYDEMEECSQAAPYIEQAQVIAHQF 483  
Db 795 ALAAVCLTVGSVAIVGRIILSGKPAIIPDREALYQOFDEMEBECASLPYMDETRAIQGF 854  
Qy 484 KEKVLGLI 491  
Db 855 KEKVLGFI 862

RESULT 14  
S21337  
genome polyprotein S4 (NS3 region) - hepatitis C virus (fragment)

C;Species: hepatitis C virus  
C;Date: 20-Feb-1995 #sequence\_revision 19-Apr-1996 #text\_change 09-Jul-2004  
C;Accession: S21337  
R;Sato, A.  
submitted to the EMBL Data Library, April 1992  
A;Description: A sensitive serodiagnosis of hepatitis C virus infection with two cloned  
A;Reference number: S21336  
A;Accession: S21337  
A;Molecule type: Genomic RNA  
A;Residues: 1-216 <SAT>  
A;Cross-references: UNIPROT:Q68968; EMBL:X65547; NID:g59494; PID:CAA46516.1; PID:g59495  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: polyprotein

Query Match 19.3%; Score 860; DB 2; Length 216;  
Best Local Similarity 94.7%; Pred. No. 1.8e-46;  
Matches 162; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
Qy 222 TGSPIYVSTYVKFLADGCGSGGAYDIIICDECHSTDATSILGIGTVDLQDAETAGARLVVL 281  
Db 1 TGAPITYSTYVKFLADGCGSGGAYDIIICDECHSTDSILGIGTVDLQDAETAGARLVVL 60  
Qy 282 ATATPPGSVTVPHNIEEVALSTGEIPFYKKAIPLEVIKGRHLIFCHSKKKDELA 341  
Db 61 ATATPPGSVTVPHNIEEVALSNTGEIPFYKKAIPETIKGRHLIFCHSKKKDELA 120  
Qy 342 LVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDSVIDCNTC 392  
Db 121 LVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDSVIDCNTC 171

RESULT 15  
T08841  
polyprotein - douroucouli hepatitis GB virus A  
C;Species: douroucouli hepatitis GB virus A  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C;Accession: T08841  
R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
J. Gen. Virol. 79, 41-45, 1998  
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
A;Reference number: Z16486; MUID:98120818; PMID:9460920  
A;Accession: T08841  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3005 <ERK>  
A;Cross-references: EMBL:AF023425; NID:g2828599; PID:AAAC40502.1; PID:g2828600  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: polyprotein

Query Match 13.6%; Score 608; DB 2; Length 3005;  
Best Local Similarity 37.0%; Pred. No. 3.6e-29;  
Matches 155; Conservative 54; Mismatches 142; Indels 68; Gaps 14;  
Qy 22 EQKESNGPVKWSIKLGTG--LHG-----FVHFEFGDNTAGCTSAGPHNPLSTR--- 71  
Db 1002 DEREHGSIVVLGTSTTRSGTCVNGVMTTFH----GSNAR--TLAGP-VGPVNCRWWS 1054  
Qy 72 -----GCNCSIYPGCHITGHRMAWKLSAARTSG-----FVSLFAP-- 107  
Db 1055 PSDDVAVYPLPSGASCLPCCK-----GTQSVWCIRNDGALCHGRLSKLVLDLPT 1106  
Qy 108 -----GAKQ-----NETHVTGGAARTSG--LTSLSFSGASQNIQLITSDNSSPPVVP 155  
Db 1107 ISDFRGSSGSPILCDEGHVGVMMVSVLHRGVKVTGVYVVKPWETLPKDSQVKSEAPPVPG 1166  
Qy 156 QS-FQVAHLHAPTCGSGKTKVPAAYAAQGYKVLVNPVAATLGFAYMSKAHGIDPNIR 214  
Db 1167 KTGFTAPLYLPTGSGKSTRIPLEYYTKAGHKVLVNPFIATVRAMGPFMEKLSQHPPIY 1226  
Qy 215 TGVRTI---TTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDATSILGIGTVDLQ 270  
Db 1227 CGHDTTAYSTTGSPLYCTYGRPMANPRRYLRGADIVI CDECHVTDP TSVLGMGRARLL 1286

Qy 271 AETAGARLVVLATATPPGSVTVPHNIEEVALSTTGEIPFYKKAIPLEVIKGRHLIFCH 330  
Db 1287 ARECGVRLLLFATATPPGAPLAQHESIKEVPLGVGDGEVAFYGHKLPVERYRTGRHLLFCH 1346  
Qy 331 SKKKCDELAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDSVIDC 389  
Db 1347 SKVECNRLHAALSTAGCNNAVYYIRGNEQEI--PAGDVVCATDALSTGYTGGFSTVTD 1403

Search completed: November 7, 2005, 20:11:17  
Job time : 41.1255 secg



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 106.241 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-4  
Perfect score: 4455  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRTSGKMGKGYWNP 829

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1624.5	36.5	2436	2 Q81756	Q81756 hepatitis c
2	1624.5	36.5	3011	1 POLG HCV1	P26664 h genome po
3	1620.5	36.4	3011	2 Q31FE5	Q31FE5 hepatitis c
4	1597.5	35.9	3011	2 Q36579	Q36579 hepatitis c
5	1597.5	35.9	3011	2 Q36608	Q36608 hepatitis c
6	1597.5	35.9	3011	2 Q36610	Q36610 hepatitis c
7	1597.5	35.9	3015	2 Q9PMU9	Q9PMU9 hepatitis c
8	1597.5	35.9	3015	2 Q9PMU5	Q9PMU5 hepatitis c
9	1596	35.8	3011	2 Q03463	Q03463 hepatitis c
10	1590.5	35.7	3011	2 Q36609	Q36609 hepatitis c
11	1586.5	35.6	2908	2 Q61X04	Q61X04 hepatitis c
12	1583.5	35.5	3011	2 Q9ELS8	Q9ELS8 hepatitis c
13	1574	35.3	3011	2 Q9DIT6	Q9DIT6 hepatitis c
14	1559.5	35.0	3011	1 POLG HCVH	P27958 h genome po
15	1559	35.0	3010	2 Q9DTE8	Q9DTE8 hepatitis c
16	1546	34.7	3010	2 Q68788	Q68788 hepatitis c
17	1546	34.7	3010	2 Q81757	Q81757 hepatitis c
18	1546	34.7	3013	2 Q6J6P5	Q6J6P5 hepatitis c
19	1545	34.7	3010	2 Q9Q1Y1	Q9Q1Y1 hepatitis c
20	1544.5	34.7	3010	2 Q9J3G4	Q9J3G4 hepatitis c
21	1544.5	34.7	3010	2 Q9J3H0	Q9J3H0 hepatitis c
22	1544	34.7	3010	2 Q9J3G5	Q9J3G5 hepatitis c
23	1542.5	34.6	3010	2 Q9DTE7	Q9DTE7 hepatitis c
24	1542.5	34.6	3010	2 Q9Q1X1	Q9Q1X1 hepatitis c
25	1542.5	34.6	3010	2 Q9Q1X2	Q9Q1X2 hepatitis c
26	1541.5	34.6	3014	2 Q9DTE0	Q9DTE0 hepatitis c
27	1541	34.6	3010	2 Q9Q1Y2	Q9Q1Y2 hepatitis c
28	1539.5	34.6	3010	2 Q9WMX2	Q9WMX2 hepatitis c
29	1539	34.5	1984	2 Q7T4V8	Q7T4V8 hepatitis c
30	1538.5	34.5	3010	2 Q9J3G6	Q9J3G6 hepatitis c
31	1538.5	34.5	3010	2 Q9J3H5	Q9J3H5 hepatitis c

RESULT 1

ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein (Fragment).			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M32084; AAA45677.1; -			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P27958; 1A1V.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0019028; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008235; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0006350; P:transcription; IEA.			
DR	GO; GO:0019079; P:viral genome replication; IEA.			
DR	GO; GO:0019087; P:viral transformation; IEA.			
DR	InterPro; IPR000345; CytC_heme_BS.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002531; HCV NS1.			
DR	InterPro; IPR000745; HCV NS4a.			
DR	InterPro; IPR001490; HCV NS4b.			
DR	InterPro; IPR002868; HCV NS5a.			
DR	InterPro; IPR002166; HCV RdRp.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR009003; Pept_Ser_Cys.			
DR	InterPro; IPR002518; Pept_U39_HCV NS2.			
DR	InterPro; IPR007095; RNA_pol_DS_PS.			
DR	InterPro; IPR007094; RNA_pol_PSVir.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	Pfam; PF01538; HCV NS2; 1.			
DR	Pfam; PF02907; HCV NS3; 1.			
DR	Pfam; PF01006; HCV NS4a; 1.			
DR	Pfam; PF01001; HCV NS4b; 1.			
DR	Pfam; PF01506; HCV NS5a; 1.			

Q9Q1Y3 hepatitis c  
Q8DTE9 hepatitis c  
Q9Q1Y4 hepatitis c  
Q9Q1Y9 hepatitis c  
Q9Q1Z0 hepatitis c  
Q8DTE5 hepatitis c  
Q9J3G8 hepatitis c  
Q8Q1X6 hepatitis c  
Q88949 hepatitis c  
Q9J3G2 hepatitis c  
Q9Q1X7 hepatitis c  
Q8Q1X8 hepatitis c  
Q00269 h genome po  
Q807P3 hepatitis c

DR Pfam; PF00271; Helicase C; 1.  
DR Pfam; PF00998; Viral RdRP; 1.  
DR SMART; SMO0487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT NON TER 1  
FT NON TER 2436  
SQ SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;

Query Match 36.5%; Score 1624.5; DB 2; Length 2436;  
Best Local Similarity 36.1%; Pred. No. 1.7e-91;  
Matches 399; Conservative 26; Mismatches 80; Indels 601; Gaps 15;

QY 146 TDNSSPPVQSFQVAVHLHAPTGSSTKVPAAVAAAGYKVLVLPNSVAATLFGAYMSK 205  
DB 761 TDNSSPPVQSFQVAVHLHAPTGSSTKVPAAVAAAGYKVLVLPNSVAATLFGAYMSK 820

QY 206 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIG 265  
DB 821 AHGIDPNIRTVGRTITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSIILGIG 880

QY 266 TVLDOAETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVVIKGRH 325  
DB 881 TVLDOAETAGARLVVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLVVIKGRH 940

QY 326 LIFCHSKKKDELAALKVALGINAVYRGLDVSVIPTSDDVVVATDALTMTGTGDFDS 385  
DB 941 LIFCHSKKKDELAALKVALGINAVYRGLDVSVIPTSDDVVVATDALTMTGTGDFDS 1000

QY 386 VIDCNTC----- 392  
DB 1001 VIDCNTCVTQVDFSLDPTFTIETILPQDAVSRTQRRGKPGIYRFVAPGERPSG 1060

QY 393 ----- 392

DB 1061 MPDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLGPLVCQDHFEGVFTGLTHIDA 1120

QY 393 ----- 392

DB 1121 HFLSQTQSGENLPYLVAQATVCARAQAPPPSWDQWKKLIRLKLPTLHGPTPLLYRLGA 1180

QY 393 -----ACSG 396  
DB 1181 VQNEITLTPVKYIMTMSADLEVTSTWLVGVGLAALAAAYCLSTGCVIVGRVLSG 1240

QY 397 KPAILPDREVLVYREFDEMESCQHLPIYIEQGMMLAEQFKOKALGL-----SRGKPAIVPD 452  
DB 1241 KPAILPDREVLVYREFDEMESCQHLPIYIEQGMMLAEQFKOKALGLLQTSRQAE-VIAPA 1299

QY 453 KEVLVQOYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVVTP---DKEI 504  
DB 1300 VQTNWQKLETFWAKHMWNFTISGQYLAGSLTLPG--NPAILASMAFTAAVTSPITTSQTL 1357

QY 505 LYE-----AFDEMECASKAALIIEGQMAEMLKSKIQLLGL----- 541  
DB 1358 LFNILGGVAAQLAAPGAATAFVGCAGLAGAIGSVGLKGLVIDLILAGYGVAGALVAFK 1417

QY 542 -----ILRRHVGGEGAVQMMNRLIAFASRGN 568  
DB 1418 INSGEVPSTEDLVNLLPAILSPGALVGVVVCVCAAILRRHVGGEGAVQMMNRLIAFASRGN 1477

QY 569 HVSPHTHVPS----- 578  
DB 1478 HVSPHTHVPSDAAARVATILSSLTVTQLRLRLHQLWISSSECTTPCSGSLWRDWDWICEV 1537

QY 579 ----- 578

DB 1538 LSDFKTLWKAKMLPQLPGIPFVSCQGYKGWVRVDGIMHTRCHGAEITGHVKNGTWIRIV 1597

QY 579 ----- 578

DB 1598 GPRTCRNMWSTGTFPINAYTTGCTPLPAPNTYFALMRVSAEYVEIRQGVDFHYVTGTTT 1657

QY 579 -----RSRRFA-----QALPFWVARPD 594  
DB 1658 DNLKCPQVPSPEFFTELDGVRLLHRRAPPCKPILLREVSFRVGLHEYPVGSQLPCEPEPD 1717

QY 595 Y-----NPP-----LV 600  
DB 1718 VAVLTSMLTDPSSHITAEAGRRLLARGSPPSVASSASQSLKATCTATNHDSPDAELI 1777

QY 601 ET---WKK-----PDYEPVVHG-----RSRRFAQALPV 627  
DB 1778 EANLLARQMGCMNITRVESKNKVILDSFDPLVAEEDEREISVPAEILRKSRRAQALPV 1837

QY 628 WARPDYNPPLVETWKKPDPYEPVVHG 653  
DB 1838 WARPDYNPPLVETWKKPDPYEPVVHG 1863

RESULT 2  
POLG HCV1  
ID POLG HCV1 STANDARD; PRT; 3011 AA.  
AC P26664;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);  
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)  
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
OS Hepatitis C virus (isolate 1) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=91172826; PubMed=1848704;  
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
RA Gallegos C., Côté D., Medina-Selby A., Barr P.J., Weiner A.J.,  
RA Bradley D.W., Kuo G., Houghton M.;  
RA "Genetic organization and diversity of the hepatitis C virus";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
CC -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
CC hydrophobic, suggesting a possible membrane-related function. NS3  
CC and NS5 may play a role in the viral RNA replication.  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
CC precursor polyprotein, commonly with Asp or Glu in the P6  
CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC (RNA)(N).  
CC -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
CC lipoprotein envelope. The envelope consists of two proteins:  
CC protein M and glycoprotein E. The nucleocapsid is a complex of  
CC protein C and mRNA.  
CC -I- SIMILARITY: Contains 1 peptidase S29 domain.  
CC -I- SIMILARITY: Contains 1 peptidase U39 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M62321; AAA45676.1; -.  
DR PUR; A39166; GNWVC3.  
DR PDB; 1HEI; X-ray; A/B=1206-1656.  
DR PDB; 1ONB; NMR; A=1349-1507.  
DR MEROPS; S29.001; -.  
DR MEROPS; U39.001; -.

DR	InterPro; IPR001410; DEAD.	FT	STRAND	1251	1255
DR	InterPro; IPR002522; HCV capsid.	FT	STRAND	1291	1295
DR	InterPro; IPR002521; HCV core.	FT	HELIX	1296	1301
DR	InterPro; IPR002519; HCV env.	FT	TURN	1302	1303
DR	InterPro; IPR002531; HCV NS1.	FT	STRAND	1312	1315
DR	InterPro; IPR002531; HCV NS1.	FT	TURN	1316	1319
DR	InterPro; IPR000745; HCV NS4a.	FT	TURN	1323	1335
DR	InterPro; IPR001490; HCV NS4b.	FT	TURN	1336	1336
DR	InterPro; IPR002868; HCV NS5a.	FT	HELIX	1337	1339
DR	InterPro; IPR002166; HCV RdRP.	FT	TURN	1340	1340
DR	InterPro; IPR001650; Helicase C.	FT	STRAND	1343	1346
DR	InterPro; IPR009003; Pept_Ser_Cys.	FT	TURN	1352	1353
DR	InterPro; IPR002518; Pept_U35_HCV_NS2.	FT	STRAND	1363	1366
DR	InterPro; IPR004109; Peptidase S29.	FT	STRAND	1373	1373
DR	InterPro; IPR007095; RNA_pol_DS_PS.	FT	TURN	1376	1377
DR	InterPro; IPR007094; RNA_pol_PSVir.	FT	STRAND	1380	1380
DR	Pfam; PF01543; HCV capsid; 1.	FT	HELIX	1382	1385
DR	Pfam; PF01542; HCV core; 1.	FT	STRAND	1390	1393
DR	Pfam; PF01539; HCV env; 1.	FT	HELIX	1397	1408
DR	Pfam; PF01560; HCV NS1; 1.	FT	TURN	1409	1411
DR	Pfam; PF01538; HCV NS2; 1.	FT	STRAND	1414	1417
DR	Pfam; PF02907; HCV NS3; 1.	FT	TURN	1430	1431
DR	Pfam; PF01006; HCV NS4a; 1.	FT	STRAND	1433	1436
DR	Pfam; PF01001; HCV NS4b; 1.	FT	TURN	1438	1439
DR	Pfam; PF01506; HCV NS5a; 1.	FT	TURN	1442	1444
DR	Pfam; PF00271; Helicase C; 1.	FT	STRAND	1450	1453
DR	Pfam; PF00998; Viral RdRP; 1.	FT	STRAND	1456	1463
DR	SMART; SM00487; DEXDC; 1.	FT	STRAND	1471	1478
KW	3D-structure; ATP-binding; Coat protein; Core protein;	FT	STRAND	1481	1488
KW	Envelope protein; Glycoprotein; Helicase; Hydrolase;	FT	HELIX	1489	1490
KW	Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;	FT	TURN	1498	1501
KW	Serine protease; Transferase; Transmembrane.	FT	STRAND	1514	1526
FT	INIT_MET 1	FT	HELIX	1527	1527
FT	1	FT	TURN	1532	1544
FT	CHAIN 1	FT	HELIX	1546	1547
FT	CHAIN 116	FT	TURN	1555	1563
FT	CHAIN 192	FT	HELIX	1564	1564
FT	CHAIN 384	FT	TURN	1570	1578
FT	CHAIN 729	FT	TURN	1579	1580
FT	CHAIN 1006	FT	HELIX	1584	1597
FT	CHAIN 1007	FT	TURN	1598	1598
FT	CHAIN 1615	FT	TURN	1606	1611
FT	CHAIN 1616	FT	TURN	1614	1618
FT	CHAIN 1863	FT	STRAND	1622	1623
FT	CHAIN 2013	FT	STRAND	1627	1627
FT	CHAIN 2014	FT	STRAND	1635	1636
FT	CHAIN 347	FT	HELIX	1640	1652
FT	CHAIN 369	FT	SEQUENCE	3011	AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;
FT	ACT_SITE 1083	FT	STRAND	36.5%;	Score 1624.5; DB 1; Length 3011;
FT	ACT_SITE 1107	FT	STRAND	Best Local Similarity	36.1%; Pred. No. 2.3e-91;
FT	ACT_SITE 1165	FT	STRAND	Matches 399; Conservative	26; Mismatches 80; Indels 601; Gaps 15;
FT	NP_BIND 1230	FT	STRAND	146	TDNSSPPVPOSFOVAHLHAPTGSKSTKVAAYAAQYKVLNPNPSVAATLGFCAYSK 205
FT	NP_BIND 1237	FT	STRAND	1211	TDNSSPPVPOSFOVAHLHAPTGSKSTKVAAYAAQYKVLNPNPSVAATLGFCAYSK 1270
FT	NP_BIND 1316	FT	STRAND	206	ARGIDPNIRTVRTITGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDATSIILGIG 265
FT	NP_BIND 1319	FT	STRAND	1271	ARGIDPNIRTVRTITGSPITYSTYTGKFLADGGCGGAYDIIICDECHSDATSIILGIG 1330
FT	CARBOHYD 196	FT	STRAND	266	TVLDOAETAGARLVVLATATPPGVTVPHPNIEEVALSTGGEIIPYKAIPLVTKGGRH 325
FT	CARBOHYD 209	FT	STRAND	1331	TVLDOAETAGARLVVLATATPPGVTVPHPNIEEVALSTGGEIIPYKAIPLVTKGGRH 1390
FT	CARBOHYD 234	FT	STRAND	326	LIFCHSKKKCDLAALKVALGINAVAYRGUDSVIPTSGDVVVVATDALTMTGYTGFDS 385
FT	CARBOHYD 305	FT	STRAND	1391	LIFCHSKKKCDLAALKVALGINAVAYRGUDSVIPTSGDVVVVATDALTMTGYTGFDS 1450
FT	CARBOHYD 307	FT	STRAND	386	VICNTC-----
FT	CARBOHYD 417	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 423	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 430	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 448	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 476	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 532	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 540	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 556	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 576	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 623	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 645	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 2041	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 2077	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 2240	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 2364	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 2789	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1224	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1232	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1236	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1238	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1246	FT	STRAND	1451	VICNTC-----
FT	CARBOHYD 1247	FT	STRAND	1451	VICNTC-----

```
QY 393 ----- 392
Db 1511 MFDSSVLCYDAGCAYELTPAETTVRLRAYMNTPLPVCQDHLERWEGVFTGLTHIDA 1570
QY 393 ----- 392
Db 1571 HFLSQTQSGENLPYLVAQYATCARAQAPPPSDQWQKCLIRLKTPLHGTPLLYRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEITLTHPVTKYITMCSADLEVTSTWLVGGVLAALAYCLSTGCVVIVGRVVLGS 1690
QY 397 KPAAIIDREVLYREFDEMESCQHLPIYIEQMMLAEOFKQKALGL-----SRGCKPAIVPD 452
Db 1691 KPAAIIDREVLYREFDEMESCQHLPIYIEQMMLAEOFKQKALGLLQOTASQRAE-VIAPA 1749
QY 453 KEVLYQOYD-----EMECSSQAAPYIEQAQVIAHQFKKVLGIIDNDVVVTP-----DKEI 504
Db 1750 VQTNWQKLETFWAKHWMNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSTQL 1807
QY 505 LYB-----AFDEMECASKAALIEGQRMASMLSKIOGLG----- 541
Db 1808 LFNILGWAAQAAPGAATAFVAGLAGAAGSGLGKVLIDILAGYAGVAGALVAFK 1867
QY 542 -----ILRRHVGPGEVAVQMMNRLIAFASRGN 568
Db 1868 IMSGEVPTEDLVNLLPAILSPALVGVVCAAILRRHVGPGEVAVQMMNRLIAFASRGN 1927
QY 569 HVSPTHYVPS----- 578
Db 1928 HVSPTHYVPSDAAARVTAISSLVTQLRLRHQWISSECTTPCSGSWLRDIWDWICEV 1987
QY 579 ----- 578
Db 1988 LSDFKTLWAKLMPQLPGIPFVSCQYKGVWRVDMHTRCHGAEITHGVNKGMTWIRIV 2047
QY 579 ----- 578
Db 2048 GPRTCRNMMSGTTPINAYTTGCTPLPAPNYTALMRVSAEYVEIRQVDFHVTGTT 2107
QY 579 -----RSRFA-----QALPVWARP 594
Db 2108 DNLKPCQVPSPEFFTELDGVLHREAPPCKPLLEBEVSPVGLHBPVGSQPCPEPD 2167
QY 595 Y-----NPP-----LV 600
Db 2168 VAVLTSMLTDPSSHITAEAGRLRAGSPPSVASSASQSLKATCTANHSDPAELI 2227
QY 601 ET---WKK-----PDYEPVVVHG-----RSSRFAQALPV 627
Db 2228 EANLLWRQMGGNITRVESKVVILDSFDPFLVAEEDEREISVPAELIRKSRFAQALPV 2287
QY 628 WARPDPNPPLVETWKKPDYEPVVVHG 653
Db 2288 WARPDPNPPLVETWKKPDYEPVVVHG 2313
```

## RESULT 3

```
Q91FE5 ID Q91FE5 PRELIMINARY; PRT; 3011 AA.
AC Q91FE5;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
```

```
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
DR EMBL; AF271632; AAF81759.1; -.
DR PIR; A44150; A44150.
DR PIR; Q0804; Q0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; Q8JYS1; ICWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR016550; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept Ser Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02307; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;
```

Query Match 36.4%; Score 1620.5; DB 2; Length 3011;

Best Local Similarity 36.0%; Pred. No. 4.1e-91;

Matches 398; Conservative 27; Mismatches 80; Indels 601; Gaps 15;

QY 146 TONSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205

Db 1211 TONSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270

QY 206 AHGIDNIRNTGVRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATSILIG 265

Db 1271 AHGIDNIRNTGVRTITGSPITYSTYVKFLADGCGSGGAYDIIICDECHSTDATSILIG 1330

QY 266 TVLDQAETAGARLVLATATPPGSVTPHPNTEEVALSTTGEIPFYGKAIPLEVIKGRH 325

Db 1331 TVLDQAETAGARLVLATATPPGSVTPHPNTEEVALSTTGEIPFYGKAIPLEVIKGRH 1390

```
QY 326 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 385
DB 1391 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDS 1450
QY 386 VIDCNC-----392
DB 1451 VIDCNCVTOTVDFSLDPTTITITLPQDAVSRQRRGTRGKPGIYRFVAPGERPSG 1510
QY 393 -----392
DB 1511 MFDSSVLCECYDAGCANNWELTPAETTVRLRAYNMTPLPVCQDHLFEWEGVFTGLTHIDA 1570
QY 393 -----392
DB 1571 HFLSOTKQSGENLPYLVAQATVCARAQAPPSWDQWKCLIRLKPTLHGPTPLLYRLGA 1630
QY 393 -----ACSG 396
DB 1631 VQNEITLTHPVTKYIMTCMSADLEVVTSTWVLGGVLAALAAAYCLSTGCVVIGRVVLSG 1690
QY 397 KPAIIPREVLVREFDEMECSQHLPIEQMMLAEQFKQKALGL-----SRGGRPAIIVPD 452
DB 1691 KPAIIPREVLVQEFDEMECSQHLPIEQMMLAEQFKQKALGLLQTSRQAE-VIAPA 1749
QY 453 KEVLYQDYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEI 504
DB 1750 VQTNWQKLETFWAKHWNFTISGIQYLAGLSTLPG--NPAIASLMAFTAATAVTSPLTTSQTL 1807
QY 505 LYE-----AFDEMECASKAALIEGQRMMAELKSKIQGLLG-----541
DB 1808 LFNILGGVAAQLAAPGAATAFVGAGLAGAIGSVGLGLKVIDLIDLAGVGAGVALVAFK 1867
QY 542 -----TLRRHVGPGEAGVQWMMRLIAFASRGN 568
DB 1868 IMSEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQWMMRLIAFASRGN 1927
QY 569 HVSPTHYVPS-----578
DB 1928 HVSPTHYVPESDAAARVAILLSLTVTQLRLRLHQWISSECTTPCSGSLRDIWDWICEV 1987
QY 579 -----578
DB 1988 LSDFKTLWAKMLQPLGIPFVSCQYKGVWGDGIMHTRCHGAEITHGVKNGTMRIV 2047
QY 579 -----578
DB 2048 GPRTCRNWSGTFPINAYTTGCTPLPAPNTYFALWRVSAEYVEIRQVGFHYVTGTT 2107
QY 579 -----RSRFA-----QALPWARP 594
DB 2108 DNLKPCQVPSPFEFFTELDGVLRLHRFAPPCKPLLRREEVSPVGLHYPVGSQLPCEPDP 2167
QY 595 Y-----NPP-----LV 600
DB 2168 VAVLTSLMTPSHITAEAGRLARGSPSVASSASQLSAPSLKATCTANHDSFDAL 2227
QY 601 ET---WKK-----PDYEPVVHG-----SSRRRFAQALPV 627
DB 2228 EANILMRMGNGNTRVSEKNVILDSFDPLVAEEDEREISVPAEILKRRRFAQALPV 2287
QY 628 WARDYNPPLVETWKKPDYEPVVHG 653
DB 2288 WARDYNPPLVETWKKPDYEPVVHG 2313
```

## RESULT 4

```
ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Polyprotein.

Query Match 35.9%; Score 1597.5; DB 2; Length 3011;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

QY 146 TDNSPPVVPQSFQVAHLHAPTGSKSTKPAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
DB 1211 TDNSPPVVPQSFQVAHLHAPTGSKSTKPAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270
```

```
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "Transmission of hepatitis C by intrahepatic inoculation with
RT transcribed RNA."
RL Science 277:570-574(1997).
DR EMBL; AF009606; AAB6324.1; -
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR PDB; 1N1L; X-ray; A/B=1017-1214.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR GO; GO:000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39 HCV NS2.
DR InterPro; IPR007095; RNA_pol_Ds_P5.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327184 MW; E2E0EE809C63C1B9 CRC64;
```

```
QY 206 AHGIDNPRTGVRITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSILGIG 265
Db 1271 AHGVDNPRTGVRITTTGSPITYSTYKFLADGCGSGAYDIIICDECHSTDATSILGIG 1330
QY 266 TVLDQAEATAGARLVWLATATPPGSGTVVPHNIEBEVALSTTGEIPFYKKAIPLEVIKGRH 325
Db 1331 TVLDQAEATAGARLVWLATATPPGSGTVVSHNIEBEVALSTTGEIPFYKKAIPLEVIKGRH 1390
QY 326 LIFCHSKKKDELAALKVALGINAVAYRGIDVSVIPTSQDGVVVVATDALMTGVTGDFDS 365
Db 1391 LIFCHSKKKDELAALKVALGINAVAYRGIDVSVIPTSQDGVVVVATDALMTGVTGDFDS 1450
QY 386 VIDCNTC----- 392
Db 1451 VIDCNCVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGTGRGKPGIYRFVAPGERPSG 1510
QY 393 ----- 392
Db 1511 MFDSSVLCECYDAGCAWYELTPAETTVRLRAYMNTPLGPLVPCODHLEFWEGVFTGLTHIDA 1570
QY 393 ----- 392
Db 1571 HFLSQTQKSGENPPYLVAQYQATVCARAQAPPSPWDQMKCLIRLKPTLHGPTLLYRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEVTLTHPTIKYIMTCSADLEVTSTWVLVGGVLAALAAVCLSTGCVVIVIGRIVLSG 1690
QY 397 KPAAIIPDREVLYREFDEMESCQHLPIYIEQMMMAEQFKOKALGL-----SRGKPAIVPD 452
Db 1691 KPAAIIPDREVLYREFDEMESCQHLPIYIEQMMMAEQFKOKALGLQLQTASRQAE-VITPA 1749
QY 453 KEVLYQOYDMEBCSQAAPYIEQAQVIAHQFKENVGLIDNDQV-----VT 499
Db 1750 VQTNWOKL-EVFWAKHWNFIQGIQYLAG-----LSTPCNPAIASLMAFTAAVTSPLT 1802
QY 500 PDKEIILYE-----AFDEMEECASKAALIEEQORMAELKSIQGLLG- 541
Db 1803 TQQTLLFNILGGVAAQAAPGAATAPVAGLAGAAGTSGVLGKVLVDILAGYGAGVAGA 1862
QY 542 -----ILRRHVGPGEAGVQVWNRLLIAP 563
Db 1863 LVAFKIMSGEVPSTEDLVNLLPAISLPGALWGVVCAAILRRHVGPGEAGVQVWNRLLIAP 1922
QY 564 ASRGNHVSPTHYVPS----- 578
Db 1923 ASRGNHVSPTHYVPSDAAARVTAIISLTVTQLRLRHQWISSECTPCSGSWLRDIWD 1982
QY 579 ----- 578
Db 1983 WICEVLSDFKTLKAKLMPQLPGIPFVSCQGYGVWRGDMHTRCHCGAEITGHVKNQ 2042
QY 579 ----- 578
Db 2043 TMRIVGPTRCNMWSGTFPINAVTTGCTPLPAPNYKFWALWRVSAEYVIBRRVGDPHYV 2102
QY 579 -----RSRFA-----QALPV 599
Db 2103 SGMVTTNLIKPCQIPSPFEFTTELDGVLRLHRFAPCKPLLRBEVSVFRVGLHEYPVGSOLPC 2162
QY 590 WARPDY-----NPP----- 598
Db 2163 EPEPDVAVLTSMLTDPSSHITAAGARRLARGSPSPMASSASQLSAPSLKATCTANHDS 2222
QY 599 ---LVET---WKK-----PDYEPVPVHG-----RSSRFA 622
Db 2223 DAELIEANLLWRQEMGGNITRVESENKVWILDSFDPLVAEEDEREVSVPAAIILKSRFA 2282
QY 623 QALPVWARPDPNPLVETWKKPDYEPVPVHG 653
Db 2283 RALPVWARPDPNPLVETWKKPDYEPVPVHG 2313
```

```
RESULT 5
O36608 PRELIMINARY; PRT; 3011 AA.
AC O36608;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus strain H77.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.
OX NCBI_TaxID=63746;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77.
RX MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus
RT are infectious when directly transfected into the liver of a
RT chimpanzee.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).
DR ENBL; AF011751; AAB67036.1; -.
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P27958; 1HEI.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5b.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Peptidase_S29.
DR InterPro; IPR002518; Peptidase_S29.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR SMART; SM00487; DEXDG; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327114 MW; 0B75B6B81CB5C198 CRC64;
```

Query Match 35.9%; Score 1597.5; DB 2; Length 3011;  
Best Local Similarity 35.4%; Pred. No. 1.1e-89;  
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

QY 146 TDNSSPPVPOSFOVAHLHAPTSGSKTKVPAAYAAOGYKVLNPNPSVAATLGFAYMSK 205  
DB 1211 TDNSSPPVPOSFOVAHLHAPTSGSKTKVPAAYAAOGYKVLNPNPSVAATLGFAYMSK 1270

QY 206 AHGIDPNIRGTGRTITTYGKFLADGCGSGAYDIIICDCHSDTATSILIG 265  
DB 1271 AHGVDPNIRGTGRTITTYGKFLADGCGSGAYDIIICDCHSDTATSILIG 1330

QY 266 TVLDAQATAGARLVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLEVIKGRH 325  
DB 1331 TVLDAQATAGARLVLTATPPGVTVPHPNIEVALSTTGEIPFYKAIPLEVIKGRH 1390

QY 326 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 385  
DB 1391 LIFCHSKKCKDELAALKVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 1450

QY 386 VIDCNTC----- 392  
DB 1451 VIDCNTCWTQTVDFSLDPTFTIETTLPODAVSTQRRGRCKPGIYRFVAPGERPSG 1510

QY 393 ----- 392  
DB 1511 MFDSSVLCYDAGCAWVELTPAETTVLRAYMNTPLGVQDHLFEWGFVFTGLTHIDA 1570

QY 393 ----- 392  
DB 1571 HFLSQTKSGENFPYLVAYQATVCARQAAPPSPWDQMKLIRLKP TLHGTP TLLYRLGA 1630

QY 393 -----ACSG 396  
DB 1631 VQNEVTLTHPTIKYIMTCSADLEVTSTWLVGVGLAALAAAYCLSTGCVVIVGRIVLSG 1690

QY 397 KPAIIPREVLRYRFBDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452  
DB 1691 KPAIIPREVLRYRFBDEMECSQHLPIYEQGMMLAEQFKQKALGLQTLASRAHE-VITPA 1749

QY 453 KEVLVYQYDEMECSQAAPYIEQQAIVTAHQEKVGLINDQVV-----VT 499  
DB 1750 VQTWQKL-EVFWAKHMWNFISSGQYLAG-----LSTLPCNPAIASMAFTAATVSPLT 1802

QY 500 PDKELIYE-----AFDEMECASKAALIEEGORMAEMLSKQGLIG- 541  
DB 1803 TQQTLLFNILGCVAAQAAPGAATAFVGAGLAGAAGTSGVLGKVLVDILAGYGAVAGA 1862

QY 542 -----ILRRHVGPGEAGVQVMNRLIAF 563  
DB 1863 LVAFKIMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQVMNRLIAF 1922

QY 564 ASRGNHVSPTHYVS----- 578  
DB 1923 ASRGNHVSPTHYVSPESDAAARVTAISLSLTVTQLRLHWHISSECTTPCSCSWLRDIWD 1982

QY 579 ----- 578  
DB 1983 WICEVLDFKTLWAKLMPQLPGIPFVSCQGRGVWRGDMIMTRCHGAEITGHVKNQ 2042

QY 579 ----- 578  
DB 2043 TMRIVGPRTCRNMWSTGTPFINAYTGPCTPLPAPNYKPALWRVSAEYVEIRRVGDHFV 2102

QY 579 -----RSRRFA-----QALPV 589  
DB 2103 SGMTTDLNLCQIPSPSEFFTELDGVRHLRFPAPCKPLLRBEVSFRVGLHEYPVGSQLP 2162

QY 590 WARDPY-----NPP----- 598  
DB 2163 EPEDVAVLTSMLTDPSSHITAEAGRRRLARGSPSPSMASASQLSAPSKATCTANHDSP 2222

QY 599 ---LVET---WKK-----PDYEPVPVHG-----BSSRRFA 622  
DB 2223 DAELIEANLWROBMGNITRVESENKVVLDSDPLVAEEDEREVSVPAILRKSRRFA 2282

QY 623 QALPVWARPDPYNPPLVETWKKPDYEPVPVHG 653  
DB 2283 RALPVWARPDPYNPPLVETWKKPDYEPVPVHG 2313

RESULT 6  
O36610 PRELIMINARY; PRT; 3011 AA.  
AC O36610;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus strain H77.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.  
OX NCBI\_TaxID=63746;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H77;  
MEDLINE=97395173; PubMed=9238047; DOI=10.1073/pnaa.94.16.8738;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Transcripts from a single full-length cDNA clone of hepatitis C virus  
are infectious when directly transfected into the liver of a  
chimpanzee";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).  
DR EMBL; AF011753; AAB67038.1; --  
DR PIR; A44150; A44150.  
DR PIR; PQ0804; PQ0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSP; P27958; 1HEI  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0019033; F:ATP binding; IEA.  
DR GO; GO:000524; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0008026; F:RNA binding; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR02522; HCV capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_Ds\_P5.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.

```
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327222 MW; 293F91430A0D4067 CRC64;

Query Match          35.9%; Score 1597.5; DB 2; Length 3011;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 33; Mismatches 75; Indels 611; Gaps 15;

QY 146 TDNSSPPVPOSQVQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNSVAATLFGAYMSK 205
DB 1211 TDNSSPPVPOSQVQVAHLHAPTSGSKTKVPAAYAAQGYKVLVNSVAATLFGAYMSK 1270
QY 206 AHGIDPNIRGTGRTITGSPITYSTYTGKFLADGCGSGAYDIIICDECHSTDATSIILGIG 265
DB 1271 AHGVDNIRGTGRTITGSPITYSTYTGKFLADGCGSGAYDIIICDECHSTDATSIILGIG 1330
QY 266 TVLDQAEATAGARLVVLATATPPGCVTPVHPNIEVALSTTGEIPFYGKAIPLEVIKGRH 325
DB 1331 TVLDQAEATAGARLVVLATATPPGCVTPVHPNIEVALSTTGEIPFYGKAIPLEVIKGRH 1390
QY 326 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGYTGDGDS 385
DB 1391 LIFCHSKKKDELAALKVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGYTGDGDS 1450
QY 386 VIDCNC- 392
DB 1451 VIDCNCVTQTVDFSLDPTFTIETTLPODAVSRTOGRGTGRGKPGIYFVAPGERPSG 1510
QY 393 - 392
DB 1511 MFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLGVPQCQDHLFEWGFGLTHIDA 1570
QY 393 - 392
DB 1571 HFLSQTQSGENPFYLVAYQATVCARAQAPPPSDQWQMKLIRLKPTLHGTPTELLVRLGA 1630
QY 393 - 396
DB 1631 VQNEVTLTHPTIKYIMTCSADLEVTVSTWVLVGGVLAALAYCLSTGCVIVGRIVLSG 1690
QY 397 KPAILPDREVLYREFDEMESCQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
DB 1691 KPAILPDREVLYQEFDEMESCQHLPIYIEQGMMLAEQFKQKALGLLQTASRHAEE-VITPA 1749
QY 453 KEVLVQYDEMESCQAAPYIEQAQVIAHQFKEKVLGLINDQVV-----VT 499
DB 1750 VQTNWQKL-EVFWAKHWNMFISGIQYLAG-----LSTLPGNPAIASLMAFTAAVTSPLT 1802
QY 500 PDEKILYE-----AFDEMECAKAALEEGORMAEMLKSKITQGLLG- 541
DB 1803 TGTLLFNILGGVAAQLAAPGRATAFVGAGLAGAAGSVGLGKVLVDILAGYAGVAGA 1862
QY 542 - 578
DB 1863 LVAFKIMSGVSPSTEDLVNLLPAILSPALVGVVCAAILRRHVGPGEVAVQWNRLLIAP 1922
QY 564 ASRGNHVSPTHYPS----- 578
DB 1923 ASRGNHVSPTHYPSDAAARVTAISSLVTQLLRLHQMISSECTTPCGSWLRDIWD 1982
QY 579 - 578
DB 1983 WICEVLSDFKTLKAKLMPQLPGIPFVSCQGYRGVWRGDGIMHTRCHGCAEITGHVKNQ 2042
QY 579 - 578
DB 2043 TMRIVGPRTCRNWMSGTFPINAVTGPCTPLPAPNYKFAALRVVSAEYVEIRVGDPHYV 2102
```

```
QY 579 -----RSRRFA-----QALPV 589
DB 2103 SGMTTDNLKPCQIIPSPFEFTLDGVLHRFAPPCKPLLRREVSFRVGLHEYPVGSQQLPC 2162
QY 590 WARPDP-----NPP----- 598
DB 2163 EPEPDVAVLTSLMTPDHSHTAEAGRRRLARGSPSSMASSASQSLKATCTANHDSP 2222
QY 599 ---LVET---WKK-----PDYEPVVHG-----RSSRRFA 622
DB 2223 DAELIEANLLMQEMGNTRVESENKVVILDSFDPLVAEEDEREVSVAEILKRRFA 2282
QY 623 QALPVNARDYNPPLVETWKKPDYEPVVHG 653
DB 2283 RALPVNARDYNPPLVETWKKPDYEPVVHG 2313

RESULT 7
Q9PMU9 PRELIMINARY; PRT; 3015 AA.
AC Q9PMU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;
RT "Hepatitis C virus: an infectious molecular clone of a second major
RT genotype (2a) and lack of viability of intertypic 1a and 2a
RT chimeras.";
RL Virology 262:250-263(1999).
DR EMBL; AF177039; AAF01181.1; -.
DR ENBL; AF177037; AAF01179.1; -.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P27958; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS5a.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR002129; Pyridoxal_deC.
```



DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RDRP; 1.  
DR SMART; SMO0487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
Polyprotein; Transmembrane.  
SQ SEQUENCE 3015 AA; 328084 MW; E309F6318067D6CD CRC64;  
  
Query Match 35.9%; Score 1597.5; DB 2; Length 3015;  
Best Local Similarity 35.4%; Pred. No. 1.1e-89;  
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;  
  
QY 146 TDNSPPVQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLNPNPSVAATLGFQAYMSK 205  
DB 1215 TDNSPPAVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQYKVLNPNPSVAATLGFQAYMSK 1274  
  
QY 206 AHGIDPNRTGVRITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATSIILGIG 265  
DB 1275 AHGVDPNRTGVRITTTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATSIILGIG 1334  
  
QY 266 TVLQAEATAGARLVVLAATPPGTVTPHPNIEEVALSTTGEIPIYKAIPLVETIKGGRH 325  
DB 1335 TVLQAEATAGARLVVLAATPPGTVTPHPNIEEVALSTTGEIPIYKAIPLVETIKGGRH 1394  
  
QY 326 LIFCHSKKKDELAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGYTGFDS 385  
DB 1395 LIFCHSKKKDELAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGYTGFDS 1454  
  
QY 386 VIDNTC----- 392  
DB 1455 VIDNTCVTQTQVDFSLDPTFTTTLTPQDAVSRQRRGTRGKGIYRFVAPGERPSG 1514  
  
QY 393 ----- 392  
DB 1515 MFDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLEFWGVFTGLTHIDA 1574  
  
QY 393 ----- 392  
DB 1575 HFLSQTKSGENFPYLVAYQATVCARQAQPPSPDQMKLIRLKPRTLHGPTPLLYRLGA 1634  
  
QY 393 -----ACSG 396  
DB 1635 VQNEVTLTHPTIKYIMTCSADLEVVTSTVLVGGVLAALAAAYCLSTGCWIVGRIVLSG 1694  
  
QY 397 KPAIIPDREVLVREFDEMECSCHLPYIEQGMMLAEQFKQKALGL-----SRGKGPAIVPD 452  
DB 1695 KPAIIPDREVLVREFDEMECSCHLPYIEQGMMLAEQFKQKALGLLQTASRAE-VITPA 1753  
  
QY 453 KEVLYQQYDEMECSQAAPYIEQAQVIAHQFKELVGLINDQVV-----VT 499  
DB 1754 VQTNQKL-EVFWAKHWNFISSIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLT 1806  
  
QY 500 PKKEILYE-----AFDEMECSKAALIEGQMAEMKSKIOGLIG- 541  
DB 1807 TCQTLLFNILGGWAAQAAPGAATAFVGAGLAGAAGISVGLGVLDVLITAGYGAGVAGA 1866  
  
QY 542 -----ILRRHVGGEGAVQMMNRLIAF 563  
DB 1867 LVAFKIMSGEVSTEDLVNLLPAILSPGALVGVVCAAILRRHVGGEGAVQMMNRLIAF 1926  
  
QY 564 ASRGNHVSPTHYVPS----- 578

DB 1927 ASRGNHVSPTHYVPSDAAARVTAISSLTVTQLLRLLHQWISSECTTPCSGSLWDIWD 1986  
QY 579 ----- 578  
DB 1987 WICEVLSDFKTLAKALPQLPGIPFVSCQGRGVGMWGDGIMHTRCHCGAEITGHVXNG 2046  
QY 579 ----- 578  
DB 2047 TMRIVGPRTCRNWMSGTFPINAYTTGPTCLPAPNPKFALWRVSAEYVEIRRVGDFHYV 2106  
QY 579 -----RSRRFA-----QALPV 589  
DB 2107 SGMTDNLKPCQIPSPSEFFTELDGVRLLRPAPCKPLLRREVSVFRVGLHYPVGSQPLPC 2166  
QY 590 WARPDY-----NPP----- 598  
DB 2167 EPEFDVAVLTSMLTDPHSHTAAAGRRRLARGSPSPMASSSSASQLSAPSLKATCTANHDSP 2226  
QY 599 ---LVET---WKK-----PDVEPPVVG-----RSSRRA 622  
DB 2227 DAELTEANLLWRQEMGNITRVESENKVVILDSFDPLVAEEDEREVSVPAEILKRSRRA 2286  
QY 623 QALPVWARPDPYNPPLVETWKKPDYEPVVG 653  
DB 2287 RALPVWARPDPYNPPLVETWKKPDYEPVVG 2317  
  
RESULT 8  
Q9PMX5 PRELIMINARY; PRT: 3015 AA.  
AC Q9PMX5;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=111103;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99420396; PubMed=10489358; DOI=10.1006/viro.1999.9889;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Hepatitis C virus: an infectious molecular clone of a second major  
RT genotype (2a) and lack of viability of intertypic 1a and 2a  
RT chimeras";  
RL Virology 262:250-263(1999).  
DR EMBL; AF177040; AAF01182.1; -.  
DR EMBL; AF177038; AAF01180.1; -.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSP; P27958; 1HEI.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003969; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR00345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.

```
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR002129; Pyridoxal_deg.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01580; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3015 AA; 328159 MW; B7D23BC1F190663A CRC64;

Query Match 35.9%; Score 1597.5; DB 2; Length 3015;
Best Local Similarity 35.4%; Pred. No. 1.1e-89;
Matches 393; Conservative 32; Mismatches 75; Indels 611; Gaps 15;

Qy 146 TDNSPPVQSFQVHLHAPTGGKSTKVPAAAGQYKVLNPNPVAATLGFAYMSK 205
Db 1215 TDNSPPVQSFQVHLHAPTGGKSTKVPAAAGQYKVLNPNPVAATLGFAYMSK 1274

Qy 206 AHGIDPNIRGTVRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGIG 265
Db 1275 AHGVDPNIRGTVRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDATSIIGIG 1334

Qy 266 TVLDAQETAGARLVLATATPPGSVTPPHNIEEVALSTTGEIPFYKAIPLVIGKGRH 325
Db 1335 TVLDAQETAGARLVLATATPPGSVTPPHNIEEVALSTTGEIPFYKAIPLVIGKGRH 1394

Qy 326 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALTGTGTFDPS 385
Db 1395 LIFCHSKKCDLAAKLVALGINAVAYRGLDVSIVPTSGDVVVVSTDALMTGTFGDFDS 1454

Qy 386 VIDNCTC----- 392
Db 1455 VIDNCTCQTQVDFSLDPTFTTETTLTPQDAVSRQRTGRGKPGIYRFVARGERPSSG 1514

Qy 393 ----- 392
Db 1515 MFDSSVLCEYDAGCANYELTPAETTVRLRAYMNTPLPVQDHLFPWEGVFTGLTHIDA 1574

Qy 393 ----- 392
Db 1575 HFLSQTKQSGENFPYVAYQATVCARAQAPPPSWDQMKLIRLKLPTLHGPTPLLYRLGA 1634

Qy 393 -----ACSG 396
Db 1635 VQNEVTLTHPTIKYIMTCMSADLEVTSTWVLVGGVLAALAAAYCLSTGCVVIVGRIVLSG 1694

Qy 397 KPAIIPDREVLRYREFDEMESCQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
Db 1695 KPAIIPDREVLRYQFDEMESCQHLPIYIEQGMMLAEQFKQKALGLLTQASRHAE-VITPA 1753

Qy 453 KEVLVQYQYDEMESCQAQPIYEQAOVTAHQFEKVLGLINDQVV-----VT 499
Db 1754 VQTNWQKL-EVFWAKHWMNFISGIQYLAG-----LSTLPGNPAIASLMAFTAAVTSPLT 1806
```

```
Qy 500 PDKEILYE-----AFDEMECASKAALIEEGORMAEMLKSKIQGLLG- 541
Db 1807 TGQTLLFNILGHWAAQLAAPGAATFVGAGLAAIGSVGLKVLVDILAGYAGVAGA 1866

Qy 542 -----ILRRHVGPGEGAVQVMNRLIAF 563
Db 1867 LVAFKIMSGEVPSTEDLVNLLPALSPGALVGVVCAAILRRHVGPGEGAVQVMNRLIAF 1926

Qy 564 ASRGNHVSPTHYVPS----- 578
Db 1927 ASRGNHVSPTHYVPSDAAARVTAISSLVTQTLRLRLHQMVISSECTTPCGSWLRDIWD 1986

Qy 579 ----- 578
Db 1987 WICEVLSDFKTLKAKLMPOLPGIPFVSCQRGYGVWRGDGIMHTRCHCAEITGHVKNQ 2046

Qy 579 ----- 578
Db 2047 TMRIVGPRTCRNMWSGTFPINAVTTGCTPLPAPNVKFWALRVSAEYVEIRRVGDFHYV 2106

Qy 579 -----RSRRFA-----QALPV 589
Db 2107 SGMTTDLNLCPCQIPSPPEFTELDGVRLLHRAFPCPKLLREEVSFRVGLHEYPVGSQPLC 2166

Qy 590 WARPDPY-----NPP----- 598
Db 2167 EPEPDVAVLTSMLTDPHSHTAEAGRRLARGSPSMASASSQLSAPSLKATCTANHDS 2226

Qy 599 ---LVET---WKK-----PYEPPVVG-----RSSRRFA 622
Db 2227 DAELEIANLLRQMGNGNITRVESKNKVLDSFDPLVAEEDREVSVPAEILRKSRFA 2286

Qy 623 QALPVWARPDPNPLVETKKKPDYEPVVG 653
Db 2287 RALPVWARPDPNPLVETKKKPDYEPVVG 2317

RESULT 9
Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Polypotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9204440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishihiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91013116; PubMed=2170712;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H., Tsuda F., Miyakawa Y., Mayumi M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177(1990).
```



Db	2303 KPNVPPVHG 2313	DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
		KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
		KW	Polyprotein; Transmembrane.
		SQ	SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA985DC CRC64;
RESULT 10			
O36609			Query Match 35.7%; Score 1590.5; DB 2; Length 3011;
AD	O36609 PRELIMINARY; PRT; 3011 AA.		Best Local Similarity 35.3%; Pred. No. 2.9e-89;
IC	O36609		Matches 392; Conservative 31; Mismatches 77; Indels 611; Gaps 15;
DT	01-JAN-1998 (T-EMBLrel. 05, Created)	Qy	146 TDNSPPVQPSQFVAHLHAPTSGSKSTKVPAAAYAGQYKVLVLPNSVAATLFGAYMSK 205
DT	01-JAN-1998 (T-EMBLrel. 05, Last sequence update)	Db	1211 TDNPPPAVQPSQFVAHLHAPTSGSKSTKVPAAAYAGQYKVLVLPNSVAATLFGAYMSK 1270
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)	Qy	206 AHGIDNIRTKVTRITITGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATSIIGIG 265
DE	Polyprotein.	Db	1271 AHGVDNIRTKVTRITITGSPITVSTYKFLADGGCGGAYDIIICDECHSTDATSIIGIG 1330
OS	Hepatitis C virus strain H77.	Qy	266 TVLDQAEATAGARLVLTATATPPGSVTVPHNPBEEVALSTTGEIPFYGKAIPLEVIKGRH 325
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	Db	1331 TVLDQAEATAGARLVLTATATPPGSVTVPHNPBEEVALSTTGEIPFYGKAIPLEVIKGRH 1390
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.	Qy	326 LIFCHSKKCDLAALKVALGINAVAYRGLDVSVIPTSQDVVVVATDALMTCTGDFDS 385
OX	NCBI_TaxID=63746;	Db	1391 LIFCHSKKCDLAALKVALGINAVAYRGLDVSVIPTSQDVVVVATDALMTCTGDFDS 1450
RN	[1]	Qy	386 VIDCNTC----- 392
RP	SEQUENCE FROM N.A.	Db	1451 VIDCNTCVTQVDFSLDPTFTIETTLTPODAVSRTORRGTRGKPGIYRFVAPGERPSG 1510
RC	STRAIN=H77;	Qy	393 ----- 392
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;	Db	1511 MFDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCQDHLFEWEGVFTGLTHIDA 1570
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;	Qy	393 ----- 392
RT	"Transcripts from a single full-length cDNA clone of hepatitis C virus	Db	1571 HFLSOTKQSGENPYLVAYQATVCARAQAPPSQDQWQKCLIRLKTPLHGTPLLLRLGA 1630
RT	are infectious when directly transfected into the liver of a	Qy	393 ----- 396
RT	chimpanzee.";	Db	1631 VQNEVTLTHPTIKYIMTCSADLEVTSTWLVGGVLAALAAVCLSTGCVVIVGRIVLSG 1690
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).	Qy	397 KPAAIPDREVLRYEFDMEBECQHLPIYIEGMMLEAFQKQALGL-----SRGKPAIVPD 452
DR	EMBL; AF011752; AAB67037.1; -.	Db	1691 KPAAIPDREVLRYEFDMEBECQHLPIYIEGMMLEAFQKQALGLQATSRHAE-VITPA 1749
DR	PIR; A41150; A44150.	Qy	453 KEVLYQYDEMBECSQAAPYIEQAQVIAHQFKEKVLGLINDQV-----VT 499
DR	PIR; PQ0804; PQ0804.	Db	1750 VQTNWQKL-EVFWAKHWNPFISGIVLAG-----LSTLPGNPAIASLMAFTAAVTSPLT 1802
DR	PIR; PS0326; PS0326.	Qy	500 PDKEILYE-----AFDEMEECASKAALIEEQORMAEMLKSKIQGLLG- 541
DR	PIR; PS0327; PS0327.	Db	1803 TGTQLLFNLTGGWVAQAAPGAATAFVGAGLAGAIGSVGLCKVLVDILAGYGAGVAGA 1862
DR	PIR; PS0328; PS0328.	Qy	542 ----- 563
DR	HSP; P26664; IHEI.	Db	1863 LVAFKIMSGEVPSTEDLVNLLPAISPALVGVWCAAILRRHVGPGEAGVQNMNRLIAF 1922
DR	GO; GO:0019028; C:viral capsid; IEA.	Qy	564 ASGNHVSPTHYVPS----- 578
DR	GO; GO:0019031; C:viral envelope; IEA.	Db	1923 ASGNHVSPTHYVPSDVAARVTAISSLTVTOLLRLHQWISSECTTPCGSGWLRIWD 1982
DR	GO; GO:0005524; F:ATP binding; IEA.	Qy	579 ----- 578
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.	Db	1983 WICEVLSDFKTLKAKLMPQLPGIPFVSCORGVGVWRGDGIMHTRCHCGAEITGHVNG 2042
DR	GO; GO:0003723; F:RNA binding; IEA.	Qy	579 ----- 578
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.	Db	2043 AMRIVGPTRCRNMWSGTFLLINAYTTGCTPLPAPNYKFAWRVSAABEYVEIRRVGDFHYI 2102
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.	Qy	579 ----- 589
DR	GO; GO:0005508; F:proteolysis and peptidolysis; IEA.	Db	2103 SGMNTNLKCPQIPSPPEFTFELDGLRRLHRFAPCPKPLLEEVSFRVGLHEYPVGSQLP 2162
DR	GO; GO:0006350; P:transcription; IEA.		
DR	GO; GO:0019079; P:viral genome replication; IEA.		
DR	InterPro; IPR000345; CytC_heme_BS.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR011545; DEAD/DEAH_N.		
DR	InterPro; IPR002522; HCV_capsid.		
DR	InterPro; IPR002521; HCV_core.		
DR	InterPro; IPR002519; HCV_env.		
DR	InterPro; IPR002531; HCV_NS1.		
DR	InterPro; IPR000745; HCV_NS4a.		
DR	InterPro; IPR001490; HCV_NS4b.		
DR	InterPro; IPR002868; HCV_NS5a.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR004109; Peptidase_S29.		
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSVir.		
DR	Pfam; PF01543; HCV_capsid; 1.		
DR	Pfam; PF01542; HCV_core; 1.		
DR	Pfam; PF01539; HCV_env; 1.		
DR	Pfam; PF01560; HCV_NS1; 1.		
DR	Pfam; PF01538; HCV_NS2; 1.		
DR	Pfam; PF02907; HCV_NS3; 1.		
DR	Pfam; PF01006; HCV_NS4a; 1.		
DR	Pfam; PF01001; HCV_NS4b; 1.		
DR	Pfam; PF01506; HCV_NS5a; 1.		
DR	Pfam; PF00271; Helicase_C; 1.		
DR	Pfam; PF00998; Viral_RdRP; 1.		
DR	SMART; SM00487; DEXDC; 1.		

```
QY 590 WARPDPY-----NPP-----598
Db 2163 EPEPDVAVLTMLTDPHSHTAEAGRLARGSPSPMASSASQSLASPLKATCATNHDSP 2222
QY 599 ---LIVET---WKK-----PDYEPVPHG-----RSSRFA 622
Db 2223 DAELEIANLWROBMGNITRVSEKNKVLDSFDPLVASEDEREVSVPAAEILKRSRFA 2282
QY 623 QALPVWARPDPYNPLVETKKPKDYEPVPHG 653
Db 2283 PALPVWARPDPYNPLVETKKPKDYEPVPHG 2313

RESULT 11
Q6IX04
ID Q6IX04 PRELIMINARY; PRT; 2908 AA.
AC Q6IX04; 2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1a.
RA Brann T.W., Kottill S., Polis M., Imamichi T.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY615798; AAT44836.1; -.
DR HSP; P27958; 1A1V.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0019031; C: viral envelope; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0006350; F: ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F: hydrolase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0008268; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0005198; F: serine-type peptidase activity; IEA.
DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P: transcription; IEA.
DR GO; GO:0019079; P: viral genome replication; IEA.
DR GO; GO:0019087; P: viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRp.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39 HCV NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase C; 1.
DR Pfam; PF00998; Viral RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
```

```
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.
FT NON TER 2908
SQ SEQUENCE 2908 AA; 315737 MW; BF5A4BC591498A4F CRC64;

Query Match 35.6%; Score 1586.5; DB 2; Length 2908;
Best Local Similarity 35.0%; Pred. No. 5e-89;
Matches 387; Conservative 32; Mismatches 86; Indels 601; Gaps 15;

QY 146 TDNSSPPVPSQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLGFAYMSK 205
Db 1211 TDNSSPPVPSQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLVLPNSVAATLGFAYMSK 1270
QY 206 AHGIDPNIRTVGRTITTTGSPITTYTGKFLADGCGSGAYDIIICDECHSDATSIILGIG 265
Db 1271 AHGVDPNIRTVGRTITTTGSPITTYTGKFLADGCGSGAYDIIICDECHSDATSIILGIG 1330
QY 266 TVLDQAETAGARLVVLATATPPGSVTVPHNIEBEVALLSTTGEIPFYKKAIPLEVIKGRH 325
Db 1331 TVLDQAETAGARLVVLATATPPGSIIVPHNIEBEVALLSTTGEIPFYKKAIPLEVIKGRH 1390
QY 326 LIFCHSKKCKDELAALKLVGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 385
Db 1391 LIFCHSKKCKDELAALKLVGINAVAYRGLDVSVIPTSGDVVVVATDALMTGTGDFDS 1450
QY 386 VIDCNTC-----392
Db 1451 VIDCNTCVTQVDFSLDPTFTIETTLPODANVSRTQRRGTRGKPGIYRFVAPGERPSG 1510
QY 393 -----392
Db 1511 MFDSSVLCEYDAGCAWYELTPAETTVRLRAYMNTPLGVPQDHLFEWEGVFTGLTHIDA 1570
QY 393 -----392
Db 1571 HFLSQTQKSGENLPYLVAQYATVCARAQAPPPSDQMWKCLIRUKPTLHGTPPLLYRLGA 1630
QY 393 -----ACSG 396
Db 1631 VQNEVTLTHPVTKYIMTCSADLEVVTSTWLVGVLAALAAAYCLSTGCVVIGRVLVS 1690
QY 397 KPALIPREVLRYREFDEMEBCSQHLPYIEQGMMLAEQFKKALGL-----SRGKPAIVPD 452
Db 1691 KPAVIPREVLRYREFDEMEBCSQHLPYIEQGMMLAEQFKKALGLLOQTASRAE-VIAPA 1749
QY 453 KEVLYQQYD-----EMECCQAAPYIEQAQVIAHQFKEKVLGLDNDQVVVTP---DKEI 504
Db 1750 VQTNWQKLEAFWAKHMWNFIISGIYLAGLSTLPG--NPATASLMAFTAAVTSPLTTSOTL 1807
QY 505 LYE-----AFDEMEBCSKAALIEEGQMAEMLKSKIQLLG-----541
Db 1808 LFNILGSWVAQAALAAPGNAATAFVCGAGLAGAIGSVGLKVLVDILAGYGVAGALVAFK 1867
QY 542 -----ILRRHVGEQAVQMMNRLIAPASRGN 568
Db 1868 IMSGEPTTEDLVNLLPAILSPGALVGVVCAAILRRHVGEQAVQMMNRLIAPASRGN 1927
QY 569 HVSTHVPS-----578
Db 1928 HVSPHYVPESDAAAARVTAISSLTVTLRLRLHOMVSSECTTPCSGSLRDINDWICEV 1987
QY 579 -----578
Db 1988 LSDFKTLWAKLMPQLPGIIPFVSCQYRGVYRGDGINMHTCHCAEITGHVKNGTMRIV 2047
QY 579 -----578
Db 2048 GPKTCRNWMSGTFPINAYTTGCTPLPAPNTYTFALWRVSAEYVEIRRVGDFHYVTGMTA 2107
QY 579 -----RSRFA-----QALPVWARPDP 594
```

```
Db 2108 DNLKPCQVPSPBEFFTELDGVLRLHREAPPCKPCLLRDEVSFRVGLHDYPPVGSQQLPCBPEPD 2167
Qy 595 Y-----NPP-----LV 600
Db 2168 VAVLTSMLTDPHSHTAEARRRLARGSPPEASSASQSLRATCTTNNHSDPAELV 2227
Qy 601 ET---WKK-----PDYPPVVHG-----RSSRRFAQALPV 627
Db 2228 EANLLWRQMGGNITRVESKNVVLDSFDPFLVAEEDEREVSVAEILRLKRRFTPALPI 2287
Qy 628 WARPDPNPPLVETWKPDYBPPVVHG 653
Db 2288 WARPDPNPPLLEWKKPDYBPPVVHG 2313

RESULT 12
ID Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Desai S.M., Devare S., Yamaguchi J.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF290978; AAGO2099.1; -.
DR PIR; A44150; A44150.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P26664; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV capsid; 1.
DR Pfam; PF01542; HCV core; 1.
DR Pfam; PF01539; HCV env; 1.
DR Pfam; PF01560; HCV NS1; 1.
DR Pfam; PF01538; HCV NS2; 1.
```

```
DR Pfam; PF02907; HCV NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327108 MW; A6BECF5A3B3EE13F CRC64;

Query Match 35.5%; Score 1583.5; DB 2; Length 3011;
Best Local Similarity 35.0%; Pred. No. 8e-89;
Matches 387; Conservative 35; Mismatches 83; Indels 601; Gaps 15;

Qy 146 TONSSPPVPSQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 205
Db 1211 TONSSPPVPSQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSK 1270
Qy 206 AHGIDENIRTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDATSLIG 265
Db 1271 AHGVDNIRTVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTDATSLIG 1330
Qy 266 TVLDQAEAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRH 325
Db 1331 TVLDQAEAGARLVLATATPPGCVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRH 1390
Qy 326 LIFCHSKKCDLAALKVALGINAVAYRGLDVSVIPTS GDVVVVATDALMTGYTGFDS 385
Db 1391 LIFCHSKKCDLAALKVALGINAVAYRGLDVSVIPTS GDVVVVATDALMTGYTGFDP 1450
Qy 386 VIDCNTC----- 392
Db 1451 VIDCNTCVTQTVDFTSLDPTFTIETTLTLPQDAVSRTORRGTRGKPGIYRFVAPGRPSG 1510
Qy 393 ----- 392
Db 1511 MFDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPLPVCODHLEFWEVFTGLTHIDA 1570
Qy 393 ----- 392
Db 1571 HFLSOTKSGENFPYLVAYQATVCARAQAPPSPDWDMWKLRKLTPLHPTLLRLGA 1630
Qy 393 -----ACSG 396
Db 1631 VQNEITLTHPVTKYIMTCHSANPEVVTSTWLVGGVLAALAAAYCLSTGCWVIVGRIVLSC 1690
Qy 397 KPAAIPDREVLRYREFDEMEECSSQHLPIEQGMMLAEQFKQKALGL-----SRGKPAIVPD 452
Db 1691 KPAAIPDREVLRYREFDEMEECSSQHLPIEQGMMLAEQFKQKALGLLQTSRQAE-VITPA 1749
Qy 453 KEVLYQOYD-----EMEECSQAAPYIEQAQVTAHQFKEKVLGLINDQVVVTP---DKEI 504
Db 1750 VQTNQKLEAFWAKHMMNFISGTQYLAGLSLTPG--NPAIASLMAFTAAVTSPLTTSQTL 1807
Qy 505 LYE-----AFDEMEECASKAALIEEGORMAEMLKSKQGLLG----- 541
Db 1808 LFNILGGWAAQLAAPGAATAFVGAGLAGAAGSGVLGKVLVDILAGYNGVAGALVAFK 1867
Qy 542 -----ILRRHVGPGEGAVQVMNRLIAFASRGN 568
Db 1868 IMSGEVPSDEDLVNLLPAILSPCALVGVVCAAILRRHVGPGEGAVQVMNRLIAFASRGN 1927
Qy 569 HVSPTHYVPS----- 578
Db 1928 HVSPTHYVPSDAAARVTAIILNLTVTQLRLRHQWIGSECTTPCSGSLRDIWDWICEV 1987
Qy 579 ----- 578
Db 1988 LSDFKTFLKAKLMPQLPGIPFVSCQRGVRGWDGIMHTRCHCGAEITGHVKNGTMRIV 2047
Qy 579 ----- 578
```

Db 2048 GPRTCRNWSTGTFPINAYTTGCTPLPAPNYKALMRVSAEYVEIRRVGDFHYVSGWTT 2107  
Qy 579 -----RSRFA-----QALPVWARP 594  
Db 2108 DNLKPCQIPSPFEFTLGDVRLHRFAPPCKPLLRVEVSFRVGLHEYPVGSQFCEPEPD 2167  
Qy 595 Y-----NPP-----LV 600  
Db 2168 VAVLTSMUTDPSHITAEAGRLARGPPSSMASSASQLSKATCTTNHSDPAELI 2227  
Qy 601 ET---WKK-----PDYPPVHG-----RSRRAQALPV 627  
Db 2228 EANLWRQMGGNITRVESENKVVILDSFDPLVAEDEREVSPAEILRKSRQFARALPV 2287  
Qy 628 WARDYNPPLVETWKKPDYPPVHG 653  
Db 2288 WARDYNPPLIETWKEPDYPPVHG 2313  
RESULT 13  
Q9DIT6 PRELIMINARY; PRT; 3011 AA.  
ID Q9DIT6  
AC Q9DIT6  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Genomic RNA for polyprotein gene.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=111103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21014672; PubMed=11115058;  
RA Kumar U., Tuthill T., Thomas H.C., Monjardino J.;  
RT "Sequence, expression and reconstitution of an HCV genome from a  
RT British isolate derived from a single blood donation.";  
RL J. Viral Hepat. 7:459-465(2000).  
DR EMBL; AJ278830; CAC03609.1; --  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR HSSP; P27958; 1A1V.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH N.  
DR InterPro; IPR011492; Flavi\_DEAD.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002868; HCV NS5a.  
DR InterPro; IPR002166; HCV RdRp.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF07652; Flavi\_DEAD; 1.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01540; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
FT CHAIN 810 1026 non-structural protein 2.  
FT CHAIN 1027 1657 non-structural protein 3.  
FT CHAIN 1658 1711 non-structural protein 4a.  
FT CHAIN 2 191 core protein.  
FT CHAIN 1712 1972 non-structural protein 4b.  
FT CHAIN 1973 2420 non-structural protein 5a.  
FT CHAIN 2421 3011 non-structural protein 5b.  
FT CHAIN 192 383 envelop protein 1.  
FT CHAIN 384 809 envelop protein 2.  
SQ SEQUENCE 3011 AA; 327406 MW; 7B6264A74A5452D3 CRC64;  
Query Match 35.3%; Score 1574; DB 2; Length 3011;  
Best Local Similarity 33.2%; Pred. No. 3.1e-88;  
Matches 404; Conservative 39; Mismatches 121; Indels 654; Gaps 20;  
Qy 146 TDNSPPVVPQSFQVAHLHAPTSGSKSTKVPAAVAAQYKVLVLPNSVAATLGFQAYMSK 205  
Db 1211 TDNSAPPVVPQSFQVAHLHAPTSGSKSTKVPAAVAAQYKVLVLPNSVAATLGFQAYMSK 1270  
Qy 206 AHGIDPNIRTVRTTGSPTTSTYTYGKFLADGCGSGGAYDIIICDECHSTDATSILGIG 265  
Db 1271 AHGIDPNIRTVRTTGSPTTSTYTYGKFLADGCGSGGAYDIIICDECHSTDATSILGIG 1330  
Qy 266 TVLDOAETAGARLVLTATATPPGVSVPVPHNIEEVALSTTGTGIEPFYKAIPLVETKGRH 325  
Db 1331 TVLDOAETAGARLVLTATATPPGVSVPVPHNIEEVALSTTGTGIEPFYKAIPLVETKGRH 1390  
Qy 326 LIFCHSKKKCDLAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGYTGFDS 385  
Db 1391 LIFCHSKKKCDLAAKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALTMTGYTGFDS 1450  
Qy 386 VIDNTC----- 392  
Db 1451 VIDNTCVTQTVDPSLDPTFTIETTLTPQDAVSRTRGRGTGRGKPGIYRFVAPGERSG 1510  
Qy 393 ----- 392  
Db 1511 MFSAVLCECYDAGCANYELTPAETTVRLRAYMNTPGLPVQDHLFPWEGVFTGLTHIDA 1570  
Qy 393 ----- 392  
Db 1571 HFLSQTQSGENLPYLVAQATVCARAQAPPPSWNDQMKLIRLKLPTLHGSTPLLYRLGA 1630  
Qy 393 -----ACSG 396  
Db 1631 VQNEVTLTHPTVKYIMTCSADLEIVTSTWVLVGGVLAALAAYCLSTGCWVIIVGRVILSG 1690  
Qy 397 KPAITPDREVLRYRDEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKGAIVPD 452  
Db 1691 KPPIIPDQEVLYRDEMECSQHLPIYIEQGMMLAEQFKQKALALLQATASRAE-VIAPA 1749  
Qy 453 KEVLYQQYD-----EMEECSQAAPYIEQAQVIAHQFKEKVLGLINDQVVVTP---DKEI 504  
Db 1750 VQTNWQRLEAFWAKHWNFIQGIQVLAGLSTLPG--NPAIASLMAFTAATVTSPLTTSQTL 1807







```

Db 2163 EPSPDVAVLSTMTDPSHITAEAGRLARGSPSPSMASASQLSAPSLKATCTANHDSF 2222
Q9DTE8 ||
Qy 599 ---LVET---WKK-----PDYEPVPHG-----RSSRRFA 622
Db 2223 DAELEIANLRQMGNGNITRSESNKVVILDSFDPLVAEEDEREVSVPAILRKSRFA 2282
Qy 623 QALPVWARPDPYNPLVETWKKPDYEPVPHG 653
Db 2283 PALPVWARPDPYNPLVETWKKPDYEPVPHG 2313

RESULT 15
Q9DTE8 PRELIMINARY; PRT; 3010 AA.
AC Q9DTE8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC TSSUE=Serum;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RL HEPATOL. Res. 20:161-171(2001).
DR EMBL: AB049089; BAB18802.1; -
DR PIR: A61196; A61196.
DR PIR: PQ0804; PQ0804.
DR PIR: PS0329; PS0329.
DR HSP; Q8JYS1; ICWX.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO: GO:0003723; F: RNA binding; IEA.
DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F: serine-type peptidase activity; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P: transcription; IEA.
DR GO: GO:0019087; P: viral genome replication; IEA.
DR InterPro: IPR000345; CytC_heme_BS.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH N.
DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR000745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRP.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR004109; Peptidase S29.
DR InterPro: IPR009003; Pept_Ser_Cys.
DR InterPro: IPR002518; Pept_U39 HCV NS2.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV capsid.1.
DR Pfam: PF01542; HCV Core.1.
DR Pfam: PF01539; HCV env.1.
DR Pfam: PF01560; HCV NS1.1.
DR Pfam: PF01538; HCV NS2.1.

```

```

DR Pfam: PF02907; HCV NS3.1.
DR Pfam: PF01006; HCV NS4a.1.
DR Pfam: PF01001; HCV NS4b.1.
DR Pfam: PF01506; HCV NS5a.1.
DR Pfam: PF00271; Helicase C.1.
DR Pfam: PF00998; Viral_RdRP.1.
DR SMART; SMO0487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3010 AA; 327068 MW; E26F4D669A836C80 CRC64;

Query Match 35.0%; Score 1559; DB 2; Length 3010;
Best Local Similarity 32.4%; Pred. No. 2.7e-87;
Matches 395; Conservative 44; Mismatches 119; Indels 662; Gaps 17;

Qy 48 VHEFGDNTAGCTSAGPHFNPLSTRGCNCSIYPGHITGHRMAWKLGSAAARTSGFVSL--F 105
Db 1142 VRRRGDNRGSLSPRFISYLKSGSGGPLLCPSGHVVG-----IFRAAVCTRGVAKAVEF 1195
Qy 106 APGAKQNETHVTGGAARTTSGLTSLFSPGASQNIQLITSTDNSSPPVVPQSQFVAHLHA 165
Db 1196 IP-----VESMETWRSP-----VFTDNSSPPAVPQTQVAHLHA 1230
Qy 166 PTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGDINIRGTGVTITGSP 225
Db 1231 PTGSGKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGDVNPRTGVTITGAP 1290
Qy 226 ITYSTYTKFLADGCGSGGAYDIIICDECHSTDATSILGTIVLDQAEATAGARLVLATAT 285
Db 1291 ITYSTYTKFLADGCGSGGAYDIIICDECHSTSTSTILGTIVLDQAEATAGARLVLATAT 1350
Qy 286 PGGSVTPHPNIEEVALSTTGEIPFYGKAIPLVVIKGGRHILFCHSKKKCDELAALKLVAL 345
Db 1351 PGGSVTPHPNIEEVALSNTGEIPFYGKAIPLVVIKGGRHILFCHSKKKCDELAALKLAL 1410
Qy 346 GINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFSDVIDCNTCA----- 393
Db 1411 GVNNAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVTQTVDFSLDPTF 1470
Qy 394 ----- 393
Db 1471 TIETTTVPQDAVSRSQRRGRTGRGGIYRFVTPGERPSGMFDSVLCYDAGCAWYEL 1530
Qy 394 ----- 393
Db 1531 TPAETSVRLRAYLNTPLGLPVCQDHLFEFEGVFTGLTHIDAHFLSQTKQAGDHPFLVAYQ 1590
Qy 394 ----- 393
Db 1591 ATVCARAQAPPPSDQMWKCLTRKPTLHGPTPLLYRLGAVQNEVTLTHPIKFIWTCMA 1650
Qy 394 -----CSGKPAIIPDREVLYREFDEMEE 416
Db 1651 ADLEIATWLVGGVLAALAAAYCLTTGSVIVIGRIILSGRPVAVPDREVLYQEPDEMEE 1710
Qy 417 CSOHLPIYEGGMMLEAQKQKALGRSGGKPAIVDPKEVLYQYDEMEECSQAAPYIE-- 474
Db 1711 CASHLPYIEGGMQLAEQFKQKALGLLQ-----TATKQEAAPVVEK 1753
Qy 475 ----QAQVIAH-----QFKEKVLGLIDNDQV-----VTPDKELIY---- 506
Db 1754 WRGLEAFWAKHMWNFTISGIQYLAGLSTLPGNPAISLMAFTASITSPLTQHTLLFNILG 1813
Qy 507 -----EAFDEMEECASKAALIEBEGORMAEMLKSKIQGLLG----- 541
Db 1814 GWVAAQLAPPASAFVAGIAGAAVGSIGLGKVLVDILAGYAGVAGALVAFKVMGEM 1873
Qy 542 -----ILRRHVGPGEVAVQVMNRLIAFASRGNHVSPTH 574
Db 1874 PSTEDMVNLLPALFSPGALVVGVCVCAIILRRHVGPGEVAVQVMNRLIAFASRGNHVSPTH 1933
Qy 575 YVP--SRSRFRFAQALP-----VWARPDYNPLVETWKK----- 605

```

Db	1934	YVPESDAAARVTQVLSLTTITQLLRLHQMINEDCSTPCSGSWLRDVMWICTVLTDFKT	1993
Qy	606	-----	605
Db	1994	WLQSKLLPRMPGVFFSCQRCYKGVWRGDBGVMHTTTCPCGADIIGHVKNKSMRIVGPKTCS	2053
Qy	606	-----PDY-----	608
Db	2054	NTWHGTFPVNAYTTGCPCTPSPAPNYSKALMRVAAEYVEVTRVGDFHYVTGMTTDNVKCP	2113
Qy	609	-----EPPVV-----	613
Db	2114	CQVPAPFEFFTELDGVRHLRYAPACKPLLRDEVTQVGLNQYPVGSQLPCEPEPDVTVITS	2173
Qy	614	-----HG-----	615
Db	2174	MLTDPSHITAEAAKRLARGCPPSLASSASQLSAPSLKATCTTYHGSPPDADLIEANLLW	2233
Qy	616	-----RSSRRFAQALPVWARPDY	633
Db	2234	RQEMGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILKTRRRFPFPALPIWARPDY	2293
Qy	634	NPPLVETWKKDDYEPVWHG	653
Db	2294	NPPLIESWRKPDYVPPVWHG	2313

Search completed: November 7, 2005, 20:17:05  
Job time : 136.241 secs

**This Page Blank (uspto)**

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	5912	100.0	1099	5	Aau76378	Aau76378 HCV multi
2	5912	100.0	1099	6	Abg72262	Abg72262 HCV multi
3	5912	100.0	1099	8	AdL66809	AdL66809 HCV multi
4	4032	68.2	829	5	Aae18690	Aae18690 HCV multi
5	4032	68.2	829	7	AdC06769	AdC06769 Chimeric
6	4032	68.2	829	8	AdL66807	AdL66807 HCV multi
7	3829.5	64.8	1021	2	Aaw34481	Aaw34481 HCV anti
8	3829.5	64.8	1021	2	Aaw40039	Aaw40039 Fusion pr
9	3829.5	64.8	1021	5	Aae222050	Aae222050 pSOD/c200
10	3050.5	51.6	841	2	AAR68547	AAR68547 HCV prote
11	3050.5	51.6	841	6	AbO27020	AbO27020 Hepatitis
12	3050.5	51.6	841	7	AdA07875	AdA07875 HCV prote
13	3047.5	51.5	841	2	AAW01701	AAW01701 hSOD-HCV
14	3047.5	51.5	841	2	AAW46397	AAW46397 Amino aci
15	3047.5	51.5	841	2	AAW97609	AAW97609 Amino aci
16	3042.5	51.5	840	2	AAR14349	AAR14349 HCV prote
17	2809.5	49.2	2231	1	AAAP90164	AAAP90164 Peptide e
18	2809.5	49.2	2436	1	AAAP92050	AAAP92050 HCV prote
19	2809.5	49.2	2436	1	AAAP90288	AAAP90288 Peptide e
20	2809.5	49.2	2772	3	AAAB18540	AAAB18540 Protein e
21	2809.5	49.2	2772	8	ADN315976	ADN315976 HCV CDNA
22	2809.5	49.2	2955	2	AAAY14975	AAAY14975 Amino aci
23	2809.5	49.2	2955	3	AAAB18541	AAAB18541 Polyprot
24	2809.5	49.2	2955	8	AdN35978	AdN35978 HCV CDNA
25	2809.5	49.2	3011	2	AAR90931	AAR90931 Hepatitis

CC sample. The method of the invention provides a sensitive, accurate  
CC diagnostic and prognostic tool to provide adequate patient care and to  
CC prevent transmission of HCV by blood and by blood products, or by  
CC personal contact. Use of NS3/4a conformational epitope in combination  
CC with MEFA, provides a sensitive and reliable method for detecting early  
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing  
CC making problems, improving sensitivity in detecting antibodies by  
CC allowing a greater number of epitopes on a unit surface area of  
CC substrate, and improving substrate. Detection accuracy is increased and  
CC the incidence of false results is reduced because of the identification  
CC and the use of highly immunogenic HCV antigens which are present during  
CC the early stages of HCV seroconversion. The present amino acid sequence  
CC represents the multiple epitope fusion antigen (MEFA) 7.1 of the  
CC invention  
XX  
SQ Sequence 1099 AA;

Query Match 100.0%; Score 5912; DB 5; Length 1099;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATKAVCVLKGDPVQGIINFEQESNGPVKWSIKGLTEGLHGFVHVEFGDNTAGCTS 60  
Db 1 MATKAVCVLKGDPVQGIINFEQESNGPVKWSIKGLTEGLHGFVHVEFGDNTAGCTS 60

Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIIIGRTLIV 120  
Db 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIIIGRTLIV 120

Qy 121 HEKADDLGKGNBESYTKGNAGSLACGVIGIAQLNSGNCNCSYPGHITGHRMAWKLS 180  
Db 121 HEKADDLGKGNBESYTKGNAGSLACGVIGIAQLNSGNCNCSYPGHITGHRMAWKLS 180

Qy 181 AARTTSFVSLFAPGAKONETHVTGGAARTTSLTSPGASQNTQLIVDFIPVENLE 240  
Db 181 AARTTSFVSLFAPGAKONETHVTGGAARTTSLTSPGASQNTQLIVDFIPVENLE 240

Qy 241 TTWRSPVFTDNSSPPVQSFQVLAHPTGSGKSTKVPAAAYAAQGVKVLNPSVAATL 300  
Db 241 TTWRSPVFTDNSSPPVQSFQVLAHPTGSGKSTKVPAAAYAAQGVKVLNPSVAATL 300

Qy 301 GFGAYMSKAHIDNIRTVRTITGSPITVSTYTKFLADGGCGGAYDIIICDECHSTD 360  
Db 301 GFGAYMSKAHIDNIRTVRTITGSPITVSTYTKFLADGGCGGAYDIIICDECHSTD 360

Qy 361 ATSLIGTGLVLDQAEAGARLWLATATPPGSGVTVPHPNIEEVALSTTGEIPFYKAIPL 420  
Db 361 ATSLIGTGLVLDQAEAGARLWLATATPPGSGVTVPHPNIEEVALSTTGEIPFYKAIPL 420

Qy 421 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYYRGLDVSVIPTSGLDVVVVATDALMT 480  
Db 421 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYYRGLDVSVIPTSGLDVVVVATDALMT 480

Qy 481 GYTGDPSVIDNCNTVTVTDFSLDPTFTTITLPQDAVSRTORRGTGRGPGIYRFV 540  
Db 481 GYTGDPSVIDNCNTVTVTDFSLDPTFTTITLPQDAVSRTORRGTGRGPGIYRFV 540

Qy 541 APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPLVPCQDHLFEWGVF 600  
Db 541 APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTPLVPCQDHLFEWGVF 600

Qy 601 TGLTHIDAHFLSQTQKSGENLPYLVAQATVCARAQAPPPSWDMWKCLRLKPLTGHPT 660  
Db 601 TGLTHIDAHFLSQTQKSGENLPYLVAQATVCARAQAPPPSWDMWKCLRLKPLTGHPT 660

Qy 661 PLYRLGAVQNEITLTHPTVKYIMTCMSADLEVVTSSACSGKPAIIPDREVLYREFDMEE 720  
Db 661 PLYRLGAVQNEITLTHPTVKYIMTCMSADLEVVTSSACSGKPAIIPDREVLYREFDMEE 720

Qy 721 CSOHLPIVEQGMMLAEQFKKALGLSRGGKPAIIPDKEVLYQYQYDEMEECQAAPYIEQA 780  
Db 721 CSOHLPIVEQGMMLAEQFKKALGLSRGGKPAIIPDKEVLYQYQYDEMEECQAAPYIEQA 780

781 QVIAHQFKEKVLGLINDQVVVTPDKIELLYEAFDEMEECASKAALIEEGQMAEMLKSKI 840  
Db 781 QVIAHQFKEKVLGLINDQVVVTPDKIELLYEAFDEMEECASKAALIEEGQMAEMLKSKI 840

Qy 841 QGLLGLILRRHVGEGAVQMMNRLIAFASRGNVSPHYPVPSRRRPAQALPVMWARPDPYN 900  
Db 841 QGLLGLILRRHVGEGAVQMMNRLIAFASRGNVSPHYPVPSRRRPAQALPVMWARPDPYN 900

Qy 901 PPLVETWKKPDYPPVPHVGHSSRRFAQALPVMWARPDPYNPPLVETWKKPDYPPVPHVGHGRT 960  
Db 901 PPLVETWKKPDYPPVPHVGHSSRRFAQALPVMWARPDPYNPPLVETWKKPDYPPVPHVGHGRT 960

Qy 961 KRNTNRRPQDVKFPFGGQIVGRRGPPIPKARRPEGRTPAQPGYPWPPLYGNKDRRSTGKSW 1020  
Db 961 KRNTNRRPQDVKFPFGGQIVGRRGPPIPKARRPEGRTPAQPGYPWPPLYGNKDRRSTGKSW 1020

Qy 1021 GKFGYPWPRTKNTNRRPQDVKFPFGGQIVGRRGPPIPKARRPEGRTPAQPGYPWPPLYG 1080  
Db 1021 GKFGYPWPRTKNTNRRPQDVKFPFGGQIVGRRGPPIPKARRPEGRTPAQPGYPWPPLYG 1080

Qy 1081 NKDRRSTGKSWGKPGYPWP 1099  
Db 1081 NKDRRSTGKSWGKPGYPWP 1099

RESULT 2  
ABG72262  
ID ABG72262 standard; protein; 1099 AA.  
XX ABG72262;  
XX AC  
XX XX  
DT 06-MAR-2003 (first entry)  
XX  
DE HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).  
XX  
KW Immunoassay solid support; Hepatitis C Virus type-1; HCV-1; HCV-2;  
KW NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;  
KW MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;  
KW HCV infection; Hepatitis C Virus type-2; Hepatitis C Virus type-3;  
KW mutant; mutin.  
XX  
OS Hepatitis C virus type 1.  
OS Hepatitis C virus type 2.  
OS Hepatitis C virus type 3.  
OS Synthetic.  
OS Chimeric.  
XX  
FH Key Location/Qualifiers  
FT Region 1. .156  
FT /note= "Correspond to amino acids 1-156 of HCV-1 hSOD  
FT superoxide dismutase)"  
FT Region 159. .176  
FT /note= "Correspond to amino acids 303-320 of HCV-1 E1"  
FT Region 179. .199  
FT /note= "Correspond to consensus sequence of amino acids  
FT 390-410 of HCV-1 E2 HVR"  
FT Region 200. .230  
FT /note= "Correspond to consensus sequence of amino acids  
FT 384-414 of HCV-1 and HCV-2 E2 HVR"  
FT Region 231. .696  
FT /note= "Correspond to amino acids 1193-1658 of HCV-1  
FT helicase"  
FT Region 699. .745  
FT /note= "Correspond to amino acids 1689-1735 of HCV-1 5-1-  
FT 1 epitope"  
FT Region 748. .794  
FT /note= "Correspond to amino acids 1689-1735 of HCV-3 5-1-  
FT 1 epitope"  
FT Region 797. .843  
FT /note= "Correspond to amino acids 1689-1735 of HCV-2 5-1-  
FT 1 epitope"  
FT Region 846. .881  
FT /note= "Correspond to amino acids 1901-1936 of HCV-1

FT polypeptide C100"  
FT 884...919  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT 922...957  
FT /note= "Correspond to amino acids 2278-2313 of HCV-1 NS5  
FT region"  
FT 958...1028  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
FT 1029...1099  
FT /note= "Correspond to core region antigenic determinants  
FT from amino acids 9-32, 39-42 and 64-88 of HCV-1 and amino  
FT acids 67-84 of HCV-2"  
XX US2002146685-A1.  
XX  
XX 10-OCT-2002.  
XX  
XX 14-JUN-2001; 2001US-00881654.  
XX  
XX 15-JUN-2000; 2000US-0212082P.  
XX 02-APR-2001; 2001US-0280811P.  
XX 02-APR-2001; 2001US-0280867P.  
XX  
XX (CHIE/) CHIEN D Y.  
XX (ARCA/) ARCANGEL P.  
XX (TAND/) TANDESKE L.  
XX (GEOR/) GEORGE-NASCIMENTO C.  
XX (COIT/) COIT D.  
XX (MEDI/) MEDINA-SELBY A.  
XX  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
XX Medina-Selby A;  
XX  
XX WPI; 2003-147573/14.  
XX N-PSDB; ABX14411.  
XX  
XX Immunocassay solid support for detecting Hepatitis C Virus infection in  
XX biological samples, comprises Hepatitis C Virus conformational epitope  
XX and multiple epitope fusion antigen.  
XX  
XX Claim 25; Fig 5A-5F; 45pp; English.  
XX  
XX The present invention relates to immunocassays comprising Hepatitis C  
XX Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion  
XX antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or the  
XX multiple epitope fusion antigen react with anti-HCV antibodies present in  
XX a biological sample from an HCV-infected individual. The immunocassays  
XX methods of the invention are useful for detecting HCV infection in a  
XX biological sample. The inventive immunocassay solid support provides a  
XX sensitive and reliable method for detecting early HCV seroconversion. The  
XX assays can detect HCV infection caused by any six known genotypes of HCV.  
XX The use of the multiple epitope fusion proteins decreases masking  
XX problems, improves sensitivity in detecting antibodies by allowing a  
XX greater number of epitopes on a unit area of substrate, and improves  
XX selectivity. The present sequence represents HCV multiple epitope fusion  
XX antigen 7.1 (MEFA 7.1), a mutant HCV polypeptide derived from various  
XX regions of HCV type 1, 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide  
XX sequences  
XX  
XX Sequence 1099 AA;  
XX  
XX Query Match 100.0%; Score 5912; DB 6; Length 1099;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFVHVEFGDNTAGCTS 60  
XX  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFVHVEFGDNTAGCTS 60  
XX  
XX 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLSDHCHIIIGRTLIV 120

RESULT 3

ADL66809  
 ID ADL66809 standard; protein; 1099 AA.  
 AC ADL66809;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE HCV multiple epitope fusion antigen 7.1 (MEFA 7.1) polypeptide.  
 XX  
 XX HCV; MEFA 7.1; HCV antigen; HCV polypeptide;  
 KW multiple epitope fusion antigen; MEFA; hepatitis C virus infection;  
 KW multiple epitope fusion antigen 7.1.  
 XX  
 OS Hepatitis C virus.  
 XX  
 XX W02004021871-A2.  
 PN  
 XX 18-MAR-2004.  
 PD  
 XX 08-SEP-2003; 2003WO-US028071.  
 PF  
 XX 09-SEP-2002; 2002US-0409515P.  
 PR  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Arcangel P, Chien D;  
 PI  
 XX WPI; 2004-248333/23.  
 DR N-PSDB; ADL66808.  
 XX  
 XX Detecting hepatitis C virus (HCV) infection in a biological sample by  
 PT detecting complexes formed between the HCV antibody and the antigens from  
 PT the first region of the HCV polypeptide and the multiple epitope fusion  
 PT antigen (MEFA).  
 XX  
 XX Claim 15; SEQ ID NO 6; 93pp; English.  
 PS  
 XX The invention relates to a method of detecting hepatitis C virus (HCV)  
 CC infection in a biological sample. The method comprises providing an  
 CC immunoassay solid support comprising HCV antigens bound to it, where the  
 CC HCV antigens comprise one or more isolated antigens form a first region  
 CC of the HCV polypeptide, combining a biological sample with the solid  
 CC support under conditions that allow HCV antibodies, when present in the  
 CC biological sample, to bind to the one or more HCV antigens, adding to the  
 CC solid support a detectably labelled HCV multiple epitope fusion antigen  
 CC (MEFA), where the labelled MEFA comprises at least one epitope from the  
 CC same region of the HCV polypeptide as the one or more isolated antigens,  
 CC where the MEFA binds to the bound HCV antibody, and detecting complexes  
 CC formed between the HCV antibody and the one or more antigens from the  
 CC first region of the HCV polypeptide and the MEFA, if any, as an  
 CC indication of HCV infection in the biological sample. The method is  
 CC useful for detecting hepatitis C virus (HCV) infection in a biological  
 CC sample. This sequence represents the MEFA 7.1 polypeptide used in the  
 CC scope of the invention.  
 XX  
 XX Sequence 1099 AA;  
 SQ  
 Query Match 100.0%; Score 5912; DB 8; Length 1099;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 1099; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
 DB 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
 QY 61 AGPHFNPLSRKHGKDPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCIIIGRTLTV 120  
 DB 61 AGPHFNPLSRKHGKDPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCIIIGRTLTV 120  
 QY 121 HEKADDLKGKGNBESSTKTGNAGSLACGVIGIAQNLSGNCNCSTYPGHITGHRMAWKLS 180  
 DB 121 HEKADDLKGKGNBESSTKTGNAGSLACGVIGIAQNLSGNCNCSTYPGHITGHRMAWKLS 180

QY 181 AARTTSFVSLFAPGAKQKQETHVTGAAARTTSGLTSLFSPGASQNIQLIVDFIPVENLE 240  
 DB 181 AARTTSFVSLFAPGAKQKQETHVTGAAARTTSGLTSLFSPGASQNIQLIVDFIPVENLE 240  
 QY 241 TTMRSPVFTDSSPPVPPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL 300  
 DB 241 TTMRSPVFTDSSPPVPPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATL 300  
 QY 301 GFCAYMSKAHGDIPNIRTGVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTD 360  
 DB 301 GFCAYMSKAHGDIPNIRTGVRTITTTGSPITYSTYKFLADGGCGSGAYDIIICDECHSTD 360  
 QY 361 ATSLIGITVLDQAETAGARLVVLTATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIP 420  
 DB 361 ATSLIGITVLDQAETAGARLVVLTATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIP 420  
 QY 421 EVIKGGRHLIFCHSKKKCDLAAKLVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 480  
 DB 421 EVIKGGRHLIFCHSKKKCDLAAKLVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 480  
 QY 481 GYTGFDSVIDCNTCTVTQTVDFSLDPTFTTETITLPPQDAVSRRTQRRGRTGRGPGYR 540  
 DB 481 GYTGFDSVIDCNTCTVTQTVDFSLDPTFTTETITLPPQDAVSRRTQRRGRTGRGPGYR 540  
 QY 541 APGERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFMEGV 600  
 DB 541 APGERPSGMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFMEGV 600  
 QY 601 TGLTHIDAHFLSQTOSGENLPYLVAQVATVCARAQAPPPSDQMKCLIRLPTLHGPT 660  
 DB 601 TGLTHIDAHFLSQTOSGENLPYLVAQVATVCARAQAPPPSDQMKCLIRLPTLHGPT 660  
 QY 661 PLLYRLGAVQNETLTHPVTKYIMTQMSADLEVVTSACSKGPALIPDREVLYREFDEME 720  
 DB 661 PLLYRLGAVQNETLTHPVTKYIMTQMSADLEVVTSACSKGPALIPDREVLYREFDEME 720  
 QY 721 CSQHLPIYIEQMMLEAFQKQKALGLSRGKGPATVPDKEVLYQYDEMECSQAAPYIEQA 780  
 DB 721 CSQHLPIYIEQMMLEAFQKQKALGLSRGKGPATVPDKEVLYQYDEMECSQAAPYIEQA 780  
 QY 781 QVTAHQFKEKVLGLINDQVVTDPKEIILYEAPDEMECASKAALIEEGORMAEMLSKI 840  
 DB 781 QVTAHQFKEKVLGLINDQVVTDPKEIILYEAPDEMECASKAALIEEGORMAEMLSKI 840  
 QY 841 QGLLGLIRRHVGEGAVQMMNRLIAFASGRNHVSPTHVPSRSRPAQALPVWARPDPYN 900  
 DB 841 QGLLGLIRRHVGEGAVQMMNRLIAFASGRNHVSPTHVPSRSRPAQALPVWARPDPYN 900  
 QY 901 PPLVETWKPKDPYBPPVYVHGSRRRFAQALPVWARPDPYNPPLVETWKPKDPYBPPVHGRKT 960  
 DB 901 PPLVETWKPKDPYBPPVYVHGSRRRFAQALPVWARPDPYNPPLVETWKPKDPYBPPVHGRKT 960  
 QY 961 KRNTNRRPQDVKPPGGQIVGRRGPPIPKARRPEGTWAGPGYWPWLYGNKDRRSTGKSW 1020  
 DB 961 KRNTNRRPQDVKPPGGQIVGRRGPPIPKARRPEGTWAGPGYWPWLYGNKDRRSTGKSW 1020  
 QY 1021 GKPGYWPWPKRTKRNTNRRPQDVKPPGGQIVGRRGPPIPKARRPEGTWAGPGYWPWLYG 1080  
 DB 1021 GKPGYWPWPKRTKRNTNRRPQDVKPPGGQIVGRRGPPIPKARRPEGTWAGPGYWPWLYG 1080  
 QY 1081 NKDRRSTGKSWGKPGYPWP 1099  
 DB 1081 NKDRRSTGKSWGKPGYPWP 1099  
 RESULT 4  
 AAE18690  
 ID AAE18690 standard; protein; 829 AA.  
 XX  
 AC AAE18690;  
 XX  
 XX 17-MAY-2002 (first entry)  
 XX



DE Multiple epitope fusion antigen (MEFA) 12 protein.  
XX Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;  
KW HCV infection; MEFA 12 protein.  
XX Unidentified.  
OS  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 315 /note= "Encoded by ATG"  
FT Misc-difference 645 /note= "Encoded by GAG"  
FT  
FT  
XX WO200196875-A2.  
XX 20-DEC-2001.  
XX 14-JUN-2001; 2001WO-US019369.  
XX 15-JUN-2000; 2000US-0212082P.  
PR 02-APR-2001; 2001US-0280811P.  
PR 02-APR-2001; 2001US-0280867P.  
XX (CHIR ) CHIRON CORP.  
FA Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
XX Medina-Selby A;  
PI  
PI  
XX WPI; 2002-179522/23.  
DR N-PSDB; AAD29796.  
DR  
XX  
XX Immunassay solid support useful for detecting hepatitis C virus  
PT infection in a biological sample, comprises at least one of HCV anti-core  
PT antibody and HCV NS3/4a epitope, bound to the support.  
PT  
XX Disclosure; Fig 7; 87pp; English.  
PS  
XX  
XX The present invention relates to hepatitis C virus (HCV) core antigen and  
CC NS (nonstructural) 3/4a antibody combination assay that can detect both  
CC HCV antigens and antibodies present in a sample using a single solid  
CC matrix as well as immunoassay solid supports for use in the assay. The  
CC solid support is useful for detecting HCV infection in a biological  
CC sample. The present sequence is MEFA (multiple epitope fusion antigen) 12  
CC protein. This sequence is used in the exemplification of the invention  
XX  
XX Sequence 829 AA;  
Query Match 68.2%; Score 4032; DB 5; Length 829;  
Best Local Similarity 69.8%; Pred. No. 2.2e-2/5;  
Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
QY 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTGLGHGFVHEFGDNTAGCTS 60  
DB 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTGLGHGFVHEFGDNTAGCTS 60  
QY 61 AGPHFNPLSRKHGGPKDEERHVGDLGNVTADKGVADVSIEDSVISLSDGHCIIIGRTLIV 120  
DB 61 AGPHFNPLSTR-----  
QY 121 HEKADDLKGKNEESTKTGNAGSLACGVIGIAQNLSGNCNSIYPGHITGHRMAWKLS 180  
DB 72 -----GCNCSIYPGHITGHRMAWKLS 93  
QY 181 AARTTSGFVSLPAPKAKONETHVTGGAARTTSGLTSLFSPGASONIQLIVDFIPVENLE 240  
DB 94 AARTTSGFVSUFPAPKAKONETHVTGGAARTTSGLTSLFSPGASONIQLITS----- 145  
QY 241 TTMRSFVFTDNSSPPVWPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLNPSVAATL 300  
DB 146 -----TDNSSPPVWPQSFQVAHLHAPTSGKSTKVPAAAYAAQYKVLNPSVAATL 197  
QY 301 GFGAYMSKAHGDIPNIRTVTITTTGSPITYSTYVKFLADGGCSGGAYDIIICDECHSTD 360  
|||||

Db 198 GFGAYMSKAHGDIPNIRTVTITTTGSPITYSTYVKFLADGGCSGGAYDIIICDECHSTD 257  
QY 361 ATSLIGITGVLDQAETAGARLVVLTATPPGCVTVPHNIEEVALSTGEIPFYGKAIPL 420  
Db 258 ATSLIGITGVLDQAETAGARLVVLTATPPGCVTVPHNIEEVALSTGEIPFYGKAIPL 317  
QY 421 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSVIPTSQDVVVVATDALMT 480  
Db 318 EVIKGGRHLIFCHSKKKDELAALVALGINAVAYRGLDVSVIPTSQDVVVVATDALMT 377  
QY 481 GYTGFDSVIDCNCVTQTVDFTTITITLPOQAVSRTQRRGTRGKPGIYRFV 540  
Db 378 GYTGFDSVIDCNC----- 392  
QY 541 APGERPSGMFDSVLCYDAGCAWYELTPAETTVRLRAYMNTFGLPVCQDHLSEWGVF 600  
Db 393 ----- 392  
QY 601 TGLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDQMWKCLIRLKLPHGPT 660  
Db 393 ----- 392  
QY 661 PLLYRLGAVQNEITLTHPVTKYIMTCMSADLEVVTSSACSGKPAIIPDREVLRYREFDEME 720  
Db 393 -----ACSGKPAIIPDREVLRYREFDEME 416  
QY 721 CSQHLPIYIEQGMMLAEQFKKALGLSRGKPAIIPDREVLRYREFDEMEECQAAPYIRQA 780  
Db 417 CSQHLPIYIEQGMMLAEQFKKALGLSRGKPAIIPDREVLRYREFDEMEECQAAPYIRQA 476  
QY 781 QVIAHQPKVKVGLIINDQVVTDPKELTAEAFDEMEECASKAALIEBQORMAEMLKSKI 840  
Db 477 QVIAHQPKVKVGLIINDQVVTDPKELTAEAFDEMEECASKAALIEBQORMAEMLKSKI 536  
QY 841 QGLLGILRRHVGPEGAVQWNNRLIAFASRGNHVSPTHYVPSRRAFAALPVMWARPYN 900  
Db 537 QGLLGILRRHVGPEGAVQWNNRLIAFASRGNHVSPTHYVPSRRAFAALPVMWARPYN 596  
QY 901 PPLVETWKKPDYEPVHGRSSRRFAQALPVMWARPYNPLVETWKKPDYEPVHGRKT 960  
Db 597 PPLVETWKKPDYEPVHGRSSRRFAQALPVMWARPYNPLVETWKKPDYEPVHGRKT 656  
QY 961 KRNTNRPPQDVKFGGGQIVG-----RRGP-----PIPKARRPESGRTWAQPGY 1003  
Db 657 KRNTNRPPQDVKFGGGQIVG-----RRGP-----PIPKARRPESGRTWAQPGY 716  
QY 1004 PWPLYGNKDRRSTGKSGKPGYWPWPKTKRNTNRPPQDVKFGGGQIVG-----RRGP- 1056  
Db 717 PWPLYGNKDRRSTGKSGKPGYWPWPKTKRNTNRPPQDVKFGGGQIVG-----RRGP- 776  
QY 1057 -----PIPKARRPESGRTWAQPGYWPWPLYGNKDRRSTGKSGKPGYWPW 1099  
Db 777 LGVLATRKTSPIPKARRPESGRTWAQPGYWPWPLYGNKDRRSTGKSGKPGYWPW 829  
RESULT 5  
ADCO6769  
ID ADC06769 standard; protein; 829 AA.  
XX  
AC ADC06769;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Chimeric multiple epitope fusion antigen 12 protein.  
KW immunoassay solid support; HCV; NS3/4a; non-structural;  
KW non-A, non-B hepatitis; NANB; multiple epitope fusion antigen 12; MEFA12;  
KW chimeric.  
OS Chimeric.  
OS Synthetic.  
OS Unidentified.  
OS Hepatitis C virus.

OS Homo sapiens.  
XX US2002192639-A1.  
XX 19-DEC-2002.  
XX 14-JUN-2001; 2001US-00881239.  
XX 15-JUN-2000; 2000US-0212082P.  
XX 02-APR-2001; 2001US-0280811P.  
XX 02-APR-2001; 2001US-0280867P.  
XX (CHIE//) CHIEN D Y.  
XX (ARCA//) ARCANDEL P.  
XX (TAND//) TANDESKE L.  
XX (GEOR//) GEORGE-NASCIMENTO C.  
XX (COIT//) COIT D.  
XX (MEDI//) MEDINA-SELBY A.  
XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;  
XX Medina-Selby A;  
XX WPI; 2003-644609/61.  
XX N-PSDB; ADC06770.  
XX  
XX Immunoassay solid support for detecting hepatitis C virus infection in  
XX biological samples, comprises a hepatitis C virus anti-core antibody and  
XX an isolated hepatitis C virus NS3/4a epitope bound HCV anti-core  
XX antibody.  
XX Claim 45; Fig 7; 40pp; English.  
XX  
XX The invention relates to a novel immunoassay solid support comprising at  
XX least one hepatitis C virus (HCV) anti-core antibody and at least one  
XX isolated HCV NS3/4a (non-structural protein 3/4a) epitope bound thereto.  
XX The system of the invention may be useful for detecting HCV infection in  
XX a biological sample and for treating or detecting non-A, non-B hepatitis  
XX (NANB hepatitis). The current sequence is that of the chimeric multiple  
XX epitope fusion antigen 12 (MEFA12) protein of the invention.  
XX  
XX Sequence 829 AA;  
XX  
XX Query Match 68.2%; Score 4032; DB 7; Length 829;  
XX Best Local Similarity 69.8%; Pred. No. 2.2e-275;  
XX Matches 791; Conservative 1; Mismatches 3; Indels 338; Gaps 7;  
XX  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
XX 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVMGSIKGLTEGLHGFHVFHFGDNTAGCTS 60  
XX  
XX 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHCHIIIGRTLTV 120  
XX 61 AGPHFNPLSTR-----  
XX  
XX 121 HEKADDLKGGBNESTKTGNAGSLRACGVICIAQNLNGCNCISYPGHITGHRMAWKLS 180  
XX 72 -----GCNCSYPGHITGHRMAWKLS 93  
XX  
XX 181 AARTTSFVSLFAPGAKONETHVTGGAAARTTSLTSLFSPGASQNTQLIVDFIPVENLE 240  
XX 94 AARTTSFVSLFAPGAKONETHVTGGAAARTTSLTSLFSPGASQNTQLITS----- 145  
XX 241 TTMRSPVFTNSPPVQSFQVLAHLHAPTCSGKSTKVPAAQAQGVKVLNLPNSVAATL 300  
XX 146 -----TNSPPVQSFQVLAHLHAPTCSGKSTKVPAAQAQGVKVLNLPNSVAATL 197  
XX  
XX 301 GFAYMSKAHGIDNIRTVRTITGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTD 360  
XX 198 GFAYMSKAHGIDNIRTVRTITGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTD 257  
XX  
XX 361 ATSLIGTGVLDQAETAGARLVLATATPGSVTVVPHNIEEVALSTTGIPFYGKAIP 420  
XX 258 ATSLIGTGVLDQAETAGARLVLATATPGSVTVVPHNIEEVALSTTGIPFYGKAIP 317

QY 421 EVIKGGRHLIFCHSKKKCDELAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 480  
DB 318 EVIKGGRHLIFCHSKKKCDELAALVALGINAVAYYRGLDVSIVPTSGDVVVVATDALMT 377  
QY 481 GYTGFDSVIDCNTCVTQTVDFSLDPTFTIETITLPQDAVSRQRRGRGKPGYRFFV 540  
DB 378 GYTGFDSVIDCNTC----- 392  
QY 541 APGERPSGPFDSVLCEDYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEGVF 600  
DB 393 ----- 392  
QY 601 TGLTHIDAHFLSOTKSGENLPYLVAQVATVCARAQAPPSPQDMWKLIRLKLPTLHGPT 660  
DB 393 ----- 392  
QY 661 PLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSACSGKPAIIPDREVLYREFDEMEE 720  
DB 393 -----ACSGKPAIIPDREVLYREFDEMEE 416  
QY 721 CSOHLPIYIEQWMLABQFKQKALGLSRGKPAIVDPKEVLYQYDMEECSSQAAPYEQ 780  
DB 417 CSOHLPIYIEQWMLABQFKQKALGLSRGKPAIVDPKEVLYQYDMEECSSQAAPYEQ 476  
QY 781 QVIAHOFKEKVLGLINDQVVTDPKEIILYEAFADEMEECASKAALIEEGORMAEMLSKI 840  
DB 477 QVIAHOFKEKVLGLINDQVVTDPKEIILYEAFADEMEECASKAALIEEGORMAEMLSKI 536  
QY 841 QGLLILRRHVGPGEGAVQNMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPYN 900  
DB 537 QGLLILRRHVGPGEGAVQNMNRLIAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPYN 596  
QY 901 PPLVETWKKPDYPPVHGRSSRRFAQALPVWARPDPYNPLVETWKKPDYPPVHGRKT 960  
DB 597 PPLVETWKKPDYPPVHGRSSRRFAQALPVWARPDPYNPLVETWKKPDYPPVHGRKT 656  
QY 961 KRNTNRRPDVKFPGGQIVG-----RRGP-----PIPKARPEGRTWAQPGY 1003  
DB 657 KRNTNRRPDVKFPGGQIVGVLPRRGPRGLVLAIRKTSPIPKARPEGRTWAQPGY 716  
QY 1004 PWPLYGNKORRSTGKSWGKPGYPPWPRKTNTNRRPDVKFPGGQIVG-----RRGP- 1056  
DB 717 PWPLYGNKORRSTGKSWGKPGYPPWPRKTNTNRRPDVKFPGGQIVGVLPRRGPR 776  
QY 1057 -----PIPKARPEGRTWAQPGYPPWPLYGNKORRSTGKSWGKPGYPPW 1099  
DB 777 LGVLATRKTSPIPKARPEGRTWAQPGYPPWPLYGNKORRSTGKSWGKPGYPPW 829  
RESULT 6  
ADL66807  
ID ADL66807 standard; protein; 829 AA.  
XX  
XX ADL66807;  
XX AC AC  
XX XX  
DT 03-JUN-2004 (first entry)  
XX  
DE HCV multiple epitope fusion antigen 12 (MEFA 12) polypeptide.  
XX  
XX HCV; MEFA 12; HCV antigen; HCV polypeptide;  
XX multiple epitope fusion antigen; MEFA; hepatitis C virus infection;  
XX multiple epitope fusion antigen 12.  
XX  
XX Hepatitis C virus.  
XX  
XX WO2004021871-A2.  
XX PN  
XX 18-MAR-2004.  
XX PD  
XX 08-SEP-2003; 2003WO-US028071.  
XX PF  
XX 09-SEP-2002; 2002US-0409515P.  
XX PR





```

XX PN US5712087-A.
XX PD 27-JAN-1998.
XX PF 12-MAY-1995; 95US-00440519.
XX PR 04-APR-1990; 90US-00504352.
XX PR 07-JUL-1992; 92US-00910760.
XX PA (CHIR ) CHIRON CORP.
XX PI Kuo G, Houghton M, Choo Q;
XX DR WPI; 1998-1119973/11.
XX DR N-PSDB; AAV09990.
XX PT Immunoassays for hepatitis C virus antibodies - using combinations of
XX PT antigenic fragments of HCV polypeptide.
XX PS Example 6; Fig 4; 59pp; English.
XX CC This sequence represents a fusion protein constructed from the hepatitis
XX CC C virus core domain (which is situated at the carboxy terminus of the
XX CC fusion protein) and a c200 construct (a fusion of the NS3 and NS3
XX CC domains). This protein used in the construction of novel combinations of
XX CC HCV antigens that have a broader range of immunological activity than any
XX CC single HCV antigen. An example of such an antigen given in this
XX CC specification comprises a first antigen containing at least 8 amino acids
XX CC of the C domain of the HCV polypeptide and a second antigen comprising at
XX CC least 8 amino acids of the NS3 domain, the NS4 domain, the S domain or
XX CC the NS5 domain of the HCV polypeptide in the form of a fusion protein, a
XX CC physical mixture or bound to a solid matrix
XX SQ Sequence 1021 AA;

Query Match 64.8%; Score 3829.5; DB 2; Length 1021;
Best Local Similarity 67.2%; Pred. No. 5.8e-261;
Matches 784; Conservative 24; Mismatches 79; Indels 279; Gaps 17;

QY 1 MATKAVCVLKGDPVQGIINPEOKESGPKVWGSIKGLTEGLHGFHVFHFGDNTAGCTS 60
DB 1 MATKAVCVLKGDPVQGIINPEOKESGPKVWGSIKGLTEGLHGFHVFHFGDNTAGCTS 60
QY 61 AGPHNPLSRKHGPKDEERHVGDLGNVADKGVADVSIEDSVISLGDHCIIIGRTLW 120
DB 61 AGPHNPLSRKHGPKDEERHVGDLGNVADKGVADVSIEDSVISLGDHCIIIGRTLW 120
QY 121 HEKADDLGKGGNEESTKTGNAGSLACGVIGIAQLNSGCNCSYPGHITGHRMAWKLS 180
DB 121 HEKADDLGKGGNEESTKTGNAGSLACGVIGIAQLNSGCNCSYPGHITGHRMAWKLS 180
QY 181 AARTTSFVSLFAPCAKQNETHVGGAAARTTSLGTSFPGASQNTQLIVDFIPVENLE 240
DB 181 AARTTSFVSLFAPCAKQNETHVGGAAARTTSLGTSFPGASQNTQLIVDFIPVENLE 240
QY 241 TTMRSPVFTDNSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQGVKVLNPSVAATL 300
DB 171 TTMRSPVFTDNSPPVPPQSFQVAHLHAPTSGKSTKVPAAAYAAQGVKVLNPSVAATL 230
QY 301 GFAYMSKAHGIDPNIRTVRTITGSPITVSTYTGKFLADGCGSGGAYDIIICDECHSTD 360
DB 231 GFAYMSKAHGIDPNIRTVRTITGSPITVSTYTGKFLADGCGSGGAYDIIICDECHSTD 290
QY 361 ATSLIGITVLDQETAGARLVWLATATPPGCVTVPHNIEVALSTTGEIPFYGKAIPL 420
DB 291 ATSLIGITVLDQETAGARLVWLATATPPGCVTVPHNIEVALSTTGEIPFYGKAIPL 350
QY 421 EVIKGGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSIVPTSGDVVVAATDALMT 480
DB 351 EVIKGGRHLIFCHSKKKCDELAALKVALGINAVAYRGLDVSIVPTSGDVVVAATDALMT 410
QY 481 GYTGFDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTOGRTGRGKPIYRFV 540

```

```

Db 411 GYTGFDSVIDCNTCVTQTVDFSLDPTFTIETITLPODAVSRTOGRTGRGKPIYRFV 470
QY 541 APGERPSGMDSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTTGLPVCODHLEFEGVF 600
Db 471 APGERPSGMDSSVLCCEYDAGCAWYELTTPAETTVRLRAYMNTTGLPVCODHLEFEGVF 530
QY 601 TGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDMWKCLIRLKPETHGPT 660
Db 531 TGLTHIDAHFLSQTQKSGENLPYLVAQVATVCARAQAPPPSDMWKCLIRLKPETHGPT 590
QY 661 PLYRLGAVONEITLTHPTVKYIMTMSADLEVVT----- 696
Db 591 PLYRLGAVONEITLTHPTVKYIMTMSADLEVVTSTWLVGGVLAALAAAYCLSTGCWVI 650
QY 697 ----ACSGKPAIIPDREVLYREFDEMEBECQHLPIYIEQGMWLAEPQKQKALGL----SRG 748
Db 651 VGRWVLSGKPAIIPDREVLYREFDEMEBECQHLPIYIEQGMWLAEPQKQKALGLLQTSRQ 710
QY 749 GKPAIVDPKEVLYQOYD-----EMEBECQAAPVIEQAQVIAHQPKVKVLGLINDQVVVT 803
Db 711 AE-VIAPAVQTNQKLETFWAKHMMNFIISGIQVLAGLSTLPG--NPAIASLMAFTAATVS 767
QY 804 P----DKEILYE-----AFDEMEECASKAALIBEGORMAEMLKSKIQGL 843
Db 768 PLTTSQTLFLNIIILGGWVAQAAPGAATFVAGLAGAAGATGSLGKVLIDILAGYAGV 827
QY 844 LG-----ILRRHVGPGEGAVQVMNRL 864
Db 828 AGALVAFKINSGEVPSFEDLVNLLPALLSPGALVGVVCAAILRRHVGPGEGAVQVMNRL 887
QY 865 IAFASRGNHVSPTHYVPSRRRFAQALPVWARPDPNPLVETWKKPKDYPVPVHGRSRR 924
Db 888 IAFASRGNHVS-----GNSST- 904
QY 925 FAQALPVWARPDPNPLVETWKKPKDYPVPVHGRKTKRNTNRRPDQVKPGGGQIVG--- 981
Db 905 -----NP-----KPKQ-----KKNRNTNRRPDQVKPGGGQIVGVY 936
QY 982 ---RRGP-----PIPKARRPEGRTWAOPGYPMPLYGNK----- 1011
Db 937 LLPRGRPLGVRAKTRKTSERSQPRGRQPIPKARRPEGRTWAOPGYPMPLYGNEGCGWAG 996
QY 1012 ---DRRSTGKSWGKPGYPMPRKTKRN 1034
Db 997 WLLSPRGRSPSGPTD---PRRSRN 1019

RESULT 9
AAE22050
ID AAE22050 standard; protein; 1021 AA.
XX AC AAE22050;
XX DX 16-JUL-2002 (first entry)
XX TT pSOD/c200/core expression plasmid protein.
DE Hepatitis C virus; HCV; antigen; C domain; polyprotein; NS3 domain;
KW NS4 domain; S domain; NS5 domain; pSOD/c200/core plasmid.
XX OS Hepatitis C virus.
XX OS Unidentified.
XX OS Chimeric.
XX FH Location/Qualifiers
XX FT 1..154
XX FT /note= "hSOD"
XX FT 155..159
XX FT /note= "Linker region"
XX FT 160..899
XX FT /note= "HCV c200"
XX FT 900..902

```

```
FT /note= "Linker region"
FT 903..1021
FT /note= "HCV c22"
XX
PN US6312889-B1.
PD 06-NOV-2001.
XX
XX 12-MAY-1995; 95US-00440549.
XX
PR 04-APR-1990; 90US-00504352.
PR 07-JUL-1992; 92US-00910760.
XX
PA (CHIR ) CHIRON CORP.
XX
XX Houghton M, Choo Q, Kuo G;
XX WPI; 2002-040268/05.
DR N-PSDB; AAD35044.
XX
XX Combination of hepatitis C viral (HCV) antigens, useful in improved
PT immunoassay for detecting HCV antibodies.
XX
PS Example 6; Fig 4; 58pp; English.
XX
XX The invention relates to combination of hepatitis C viral (HCV) antigens
CC that have a broader range of immunological reactivity than any single HCV
CC antigen. The combinations consist of an antigen from the C domain of the
CC HCV polyprotein, and at least one additional HCV antigen from either the
CC NS3 domain, the NS4 domain, the S domain, or the NS5 domain and are in
CC the form of fusion protein, a simple physical mixture, or the individual
CC antigens commonly bound to a solid matrix. The combinations of antigens
CC provides broad range immunoassays for anti-HCV antibodies. The invention
CC therefore provides a method for detecting antibodies to HCV in a mammal
CC suspected of containing such antibodies. The present sequence is a
CC protein encoded by pSOD/c200/core expression plasmid DNA containing HCV
CC coding sequence
XX
XX Sequence 1021 AA;
XX
XX Query Match 64.8%; Score 3829.5; DB 5; Length 1021;
XX Best Local Similarity 67.2%; Pred. No. 5.8e-261;
XX Matches 784; Conservative 24; Mismatches 79; Indels 279; Gaps 17;
XX
QY 1 MATKAVCVLKGDCPVQGIINFEOKESGPKVKGSIKGLTEGLHGHVHFEGDNTAGCTS 60
DB 1 MATKAVCVLKGDCPVQGIINFEOKESGPKVKGSIKGLTEGLHGHVHFEGDNTAGCTS 60
QY 61 AGPHFNPLSRKHGKGPDEERHVGDLGNVTADKGVADVSIEDSVISLGDHICIGRTLTV 120
DB 61 AGPHFNPLSRKHGKGPDEERHVGDLGNVTADKGVADVSIEDSVISLGDHICIGRTLTV 120
QY 121 HEKADDLGKGNBESTKTGNAGSRLACGVGTGIAQLNSGNCNCIYPGHITGHRMAWKLS 180
DB 121 HEKADDLGKGNBESTKTGNAGSRLACGVGTGIAQLNEFGA----- 160
QY 181 AARTTSGFVSLPAPGAKQNETHTVGGAAARTTSLTSLFSPGASQNTLIQVDFIPVENLE 240
DB 161 -----VDFIPVENLE 170
QY 241 TTMRSVPFTDSSPPVVPQSFQVAHLHAPTGSKGSTKVPAAAYAAQGYKVLVLPNSVAATL 300
DB 171 TTMRSVPFTDSSPPVVPQSFQVAHLHAPTGSKGSTKVPAAAYAAQGYKVLVLPNSVAATL 230
QY 301 GFGAYMSKAHGDIPNIRTVRTITGSPITYSTYVGKFLADGGCGSGAYDIIICDECHSTD 360
DB 231 GFGAYMSKAHGDIPNIRTVRTITGSPITYSTYVGKFLADGGCGSGAYDIIICDECHSTD 290
QY 361 ATSLIGTGVLDQAEYTAGARLVLATATPGSVTVPHNIEEVALSTTGEIPFGKAIPL 420
DB 291 ATSLIGTGVLDQAEYTAGARLVLATATPGSVTVPHNIEEVALSTTGEIPFGKAIPL 350
QY 421 EVIKGGRHLIFCHSKKKCDLAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMT 480
```

## RESULT 10

AAR68547

ID AAR68547 standard; protein; 841 AA.

XX

AC AAR68547;

XX

DT 25-MAR-2003 (revised)

DT 17-AUG-1995 (first entry)

XX

XX HCV protease/hSOD fusion protein expression vector cflSODp600.

XX

KW Hepatitis C virus protease/hSOD fusion protein; HCV;

XX expression vector cflSODp600; viral infectivity inhibition.

XX

OS Hepatitis C virus.

XX

PH Key Location/Qualifiers

FT Peptide 1..155

XX /label= SOD leader

XX

PN US5371017-A.







```
Qy 61 AGPHFNPLSRKHGPGKDEERHVGDLGNVTADKOGVADVSIEDSVISLSDGHCIIIGRTLIV 120
Db 61 PGPHFNPLSRKHGPGKDEERHVGDLGNVTADKOGVADVSIEDSVISLSDGHCIIIGRTLIV 120
Qy 121 HEKADDLGKGNBEESTKTGNAGSLACGVIGIAQNLNSGCNCISYPGHIHTGHR----- 173
Db 121 HEKADDLGKGNBEESTKTGNAGSLACGVIGIR-----IGTYVY-NHLTPLRDWAHNGL 174
Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190
Db 175 RDLAVAVEPVFSQMETKLIITWGDATACGDIINGLPVSARRGRIILLGPADGMVSKGWR 234
Qy 191 LFAP-----GAKONETH-----VTG 205
Db 235 LLAPITAYAQOQTRGLLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCIINGVCWTYH 294
Qy 206 GAAART-----TSGLT----- 216
Db 295 GAGTRTITASPKGPVIMYTNVDQLVGPASQGRSLTPCTCGSSDLYLVRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Db 355 RRGDSRGLSPRPISYILKGSSSGGPLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPVFTDNSSPPVQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLG 301
Db 415 TMRSPVFTDNSSPPVQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLG 474
Qy 302 FGAYMSKAHGDIPNIRTCVRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDA 361
Db 475 FGAYMSKAHGDIPNIRTCVRTITGSPITYSTYCKFLADGCGSGAYDIIICDCHSTDA 534
Qy 362 TSIILGIGTVLDOAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYKAIPLE 421
Db 535 TSIILGIGTVLDOAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYKAIPLE 594
Qy 422 VIKGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVATDALMTG 481
Db 595 VIKGRHLIFCHSKKKKDELAALVALGINAVAYRGLDVSIVPTSGDVVVATDALMTG 654
Qy 482 YTGPFDSVIDCNTCTQTVDFSLDPTFTIETITLPQDAVSRTORGRGKPGIYRVA 541
Db 655 YTGPFDSVIDCNTCTQTVDFSLDPTFTIETITLPQDAVSRTORGRGKPGIYRVA 714
Qy 542 PGERPSGMFDSVLCEDYDAGCAYELTPAETTVRLRAYMNTPLPVQCQDHLFEWEGVFT 601
Db 715 PGERPPGMFDSVLCEDYDAGCAYELTPAETTVRLRAYMNTPLPVQCQDHLFEWEGVFT 774
Qy 602 GLTHIDAHFLSQTKSGENLPYLVAQATVCARQAQPPSWDQMKLIRLKLPTLHGPTP 661
Db 775 GLTHIDAHFLSQTKSGENLPYLVAQATVCARQAQPPSWDQMKLIRLKLPTLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841
```

RESULT 13  
AAW01701  
ID AAW01701 standard; protein; 841 AA.  
XX  
XX  
AC AAW01701;  
XX  
DT 17-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 03-APR-1997 (first entry)  
XX  
XX hSOD-HCV fusion protein.  
XX  
KW HCV; NS3; non-structural domain 3; protease; polyprotein; inhibitor;  
KW screen; processing; infection; treatment; probe; hepatitis C virus.  
XX  
OS Hepatitis C virus; Virus.

```
OS Homo sapiens.  
XX Chimeric.  
FH Key Location/Qualifiers  
FT Protein 156..841  
FT /label= HCV_protease  
XX  
FN US5585258-A.  
XX  
XX 17-DEC-1996.  
PD  
XX 06-DEC-1994; 94US-00350884.  
PF  
XX 04-APR-1990; 90US-00505433.  
PR  
XX 04-APR-1991; 91US-00680296.  
XX (CHIR ) CHIRON CORP.  
XX  
XX Choo Q, Kuo G, Houghton M;  
PI WPI; 1997-051175/05.  
XX N-PSDB; AAT59261.  
DR  
XX  
XX  
PT Compn. contg. hepatitis C virus NS3 domain protease and related fusion  
PT proteins - useful for screening specific inhibitors, potential antiviral  
PT agents, prepn. of antibodies and for cleaving specific poly.peptide(s).  
XX  
XX Example 4; Col 77-84; 68pp; English.  
PS  
XX  
XX Compns. comprising the hepatitis C virus (HCV) NS3 domain protease or  
CC its active truncation analogues are claimed. Also new are fusion proteins  
CC comprising the protease (or analogues) and, e.g. human superoxide (SOD),  
CC or ubiquitin. The protease is essential for polyprotein processing, and  
CC thus infectivity, in HCV. The compns. are used to screen for specific  
CC inhibitors (possibly useful as antiviral agents), to generate specific  
CC antibodies and to cleave specific polypeptides. HCV cDNA clones (AAT59250  
CC - 56 encoding AAW01686-92 resp.) were isolated from HCV genomic library  
CC using probes AAT59244-49. The clones were used in the preparation of full  
CC -length SOD-protease fusion proteins. The present sequence is encoded by  
CC vector cFISODp600 which contains a full-length HCV protease coding  
CC sequence fused to a functional hSOD leader. The resulting vector encodes  
CC amino acids 1-151 of hSOD, and amino acids 946-1630 of HCV (corresponding  
CC to 1-886 of AAW01693). (Updated on 25-MAR-2003 to correct PF field.)  
CC (Updated on 17-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 841 AA;  
Query Match 51.5%; Score 1047.5; DB 2; Length 841;  
Best Local Similarity 72.6%; Pred. No. 6.8e-206;  
Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;  
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Db 1 MATNPVCVLKGDGPVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60  
Qy 61 AGPHFNPLSRKHGPGKDEERHVGDLGNVTADKOGVADVSIEDSVISLSDGHCIIIGRTLIV 120  
Db 61 PGPHFNPLSRKHGPGKDEERHVGDLGNVTADKOGVADVSIEDSVISLSDGHCIIIGRTLIV 120  
Qy 121 HEKADDLGKGNBEESTKTGNAGSLACGVIGIAQNLNSGCNCISYPGHIHTGHR----- 173  
Db 121 HEKADDLGKGNBEESTKTGNAGSLACGVIGIR-----IGTYVY-NHLTPLRDWAHNGL 174  
Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190  
Db 175 RDLAVAVEPVFSQMETKLIITWGDATACGDIINGLPVSARRGRIILLGPADGMVSKGWR 234  
Qy 191 LFAP-----GAKONETH-----VTG 205  
Db 235 LLAPITAYAQOQTRGLLGCIITSLTGRDKNOVEGEVQIVSTAAQTFLATCIINGVCWTYH 294  
Qy 206 GAAART-----TSGLT----- 216
```

```

Db 295 GAGTRTTIATSPKGPVIQMYTNVDQLVGPASQGRSLTPTCTCGSSDLYLVTRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Dy 355 RRGDSRGSLLSPRPISYLVKSGSGPGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPVFTDNSSPPVPQSFQVAHLHAPTSGKSTKVPAAQAQGYKVLVLPNSVAATLG 301
Dy 415 TMRSPVFTDNSSPPVPQSFQVAHLHAPTSGKSTKVPAAQAQGYKVLVLPNSVAATLG 474
Qy 302 FGAYMSKAHGIDPNIRGTGVRTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 361
Dy 475 FGAYMSKAHGIDPNIRGTGVRTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 534
Qy 362 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLE 421
Dy 535 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLE 594
Qy 422 VIKGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTG 481
Dy 595 VIKGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTG 654
Qy 482 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTTITLPQDAVSRTOGRTGRKPGIYRFVA 541
Dy 655 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTTITLPQDAVSRTOGRTGRKPGIYRFVA 714
Qy 542 PGERPQMFSSVLCCEYDAGCANYELTPAETTVRLRAYMNTPGLPVCQDHLEFWEQVFT 601
Dy 715 PGERPQMFSSVLCCEYDAGCANYELTPAETTVRLRAYMNTPGLPVCQDHLEFWEQVFT 774
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMKCLIRLKLPTLHGPTP 661
Dy 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMKCLIRLKLPTLHGPTP 834
Qy 662 LLYRLGA 668
Dy 835 LLYRLGA 841

RESULT 14
AAW46397
ID AAW46397 standard; protein; 841 AA.
AC AC
XX XX
DT 27-AUG-2003 (revised)
DT 07-MAY-1998 (first entry)
XX XX
DE Amino acid sequence of the vector cf1SODp600.
XX XX
KW Protease; HCV; NS3 domain; human superoxide dismutase; fusion protein;
KW assay; activity; anti-HCV.
XX XX
OS Synthetic.
OS Hepatitis C virus.
XX XX
OS Homo sapiens.
XX XX
PN US5712145-A.
XX XX
PD 27-JAN-1998.
XX XX
XX 06-SEP-1996; 96US-00709173.
XX XX
PR 04-APR-1990; 90US-00505433.
PR 04-APR-1991; 91US-00680296.
PR 06-DEC-1994; 94US-00350884.
PR 12-MAY-1995; 95US-00440548.
XX XX
PA (CHIR ) CHIRON CORP.
XX XX
PI Choo Q, Kuo G, Houghton M;
XX XX
DR WPI; 1998-119986/11.

```

```

DR N-PSDB; AAV04993.
XX Recombinant hepatitis C virus protease - useful in screening drugs for
PT activity against hepatitis C virus.
PT Disclosure; Fig 10A-G; 68pp; English.
PS
XX
CC The present sequence represents the amino acid sequence of the vector
CC cf1SODp600. This vector contains a full length Hepatitis C virus (HCV)
CC protease coding sequence fused to a functional human superoxide dismutase
CC leader. The vector was used to express the protease fusion protein in
CC Escherichia coli. The HCV protease is believed to cleave itself from the
CC genomic polyprotein. In the absence of protease activity, the HCV
CC polyprotein should remain in its unprocessed form, and thus render the
CC virus non-infectious. Inhibitors of protease activity should also inhibit
CC viral infectivity. The protease can therefore be used for assaying
CC compounds for activity against HCV. (Updated on 27-AUG-2003 to correct OS
CC field.)
XX
SQ Sequence 841 AA;
Query Match 51.5%; Score 3047.5; DB 2; Length 841;
Best Local Similarity 72.6%; Pred. No. 6.8e-206;
Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;
Qy 1 MATKAVCLVKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60
Dy 1 MATNPVCLVKGDPVQGIINFEOKESNGPVKVGSIKGLTEGLHGFHVEFGDNTAGCTS 60
Qy 61 AGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHICIGRTLTV 120
Dy 61 PGPHFNPLSRKHGPKDEERHVGDLGNVTADKGVADVSIEDSVISLGDHICIGRTLTV 120
Qy 121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIAQNLSGNCNCSIIYPGHITGHR 173
Dy 121 HEKADDLGKGNESSTKTGNAGSRLACGVIGIR-----GTYYV-NHLTPLRDWAHNG 174
Qy 174 -----MAWKLGSAA-----RTTSGFVS----- 190
Dy 175 RDLAVAVEPVVFSOMETKLTWGAADTAACGDIINGLPVSARRGREILLGPADGMVSKGR 234
Qy 191 LFAP-----GAKQNEH-----VTG 205
Dy 235 LLAPITAYAQOTRGLLCIITSITGRDKQVEGEVQIVSTAQTFLATCIINGVCWTYYH 294
Qy 206 GAAART-----TSGLT----- 216
Dy 295 GAGTRTIATSPKGPVIQMYTNVDQLVGPASQGRSLTPTCTCGSSDLYLVTRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Dy 355 RRGDSRGSLLSPRPISYLVKSGSGPGLPCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPVFTDNSSPPVPQSFQVAHLHAPTSGKSTKVPAAQAQGYKVLVLPNSVAATLG 301
Dy 415 TMRSPVFTDNSSPPVPQSFQVAHLHAPTSGKSTKVPAAQAQGYKVLVLPNSVAATLG 474
Qy 302 FGAYMSKAHGIDPNIRGTGVRTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 361
Dy 475 FGAYMSKAHGIDPNIRGTGVRTITTTGSPITYSTYKFLADGCGSGGAYDIIICDECHSTDA 534
Qy 362 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLE 421
Dy 535 TSILGIGTVLDQAETAGARLVLTATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLE 594
Qy 422 VIKGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTG 481
Dy 595 VIKGGRHLIFCHSKKCKDELAALKVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTG 654
Qy 482 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTTITLPQDAVSRTOGRTGRKPGIYRFVA 541
Dy 655 YTGDFDSVIDNCNTCVTQTVDFSLDPTFTTITLPQDAVSRTOGRTGRKPGIYRFVA 714

```

```
Qy 542 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWGVFT 601
Db 715 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWGVFT 774
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTP 661
Db 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841

RESULT 15
AAW97609
ID AAW97609 standard; protein; 841 AA.
AC AAW97609;
XX 26-MAY-1999 (first entry)
XX Amino acid sequence of vector cf15ODp600.
XX HCV NS3 protease; truncation analog; HCV control; protease activity;
KW viral infectivity; inactive non-cleaving protease.
XX Synthetic.
OS Hepatitis C virus.
XX US5885799-A.
XX 23-MAR-1999.
XX 06-SEP-1996; 96US-00709177.
XX 04-APR-1990; 90US-00505433.
XX 04-APR-1991; 91US-00680296.
XX 06-DEC-1994; 94US-00350884.
XX 12-MAY-1995; 95US-00440548.
XX (CHIR ) CHIRON CORP.
XX Choo Q, Kuo G, Houghton M;
XX WPI; 1999-228536/19.
XX N-PSDB; AAX26398.
XX Preparation of new Hepatitis C Virus NS3 protease - useful for screening
XX for compounds which inhibit HCV infectivity.
XX Example 3; Fig 10; 71pp; English.
XX The specification describes a method for making a purified Hepatitis C
XX virus (HCV) NS3 protease or active truncation analog. If the HCV protease
XX N-terminal cleavage signal is excluded (so that self-cleavage is
XX prevented), the HCV protease remains in its unprocessed form, and renders
XX the virus noninfectious. The protease is therefore useful for assaying
XX pharmaceutical agents for control of HCV, as compounds which inhibit
XX protease activity sufficiently will also inhibit viral infectivity. An
XX inactive non-cleaving protease can be used to screen for inhibitors.
XX Recombinant expression systems can be utilised to prepare recombinant HCV
XX which can be used to produce monoclonal antibodies. The present sequence
XX was created in the course of the invention
XX Sequence 841 AA;
XX Query Match 51.5%; Score 3047.5; DB 2; Length 841;
XX Best Local Similarity 72.6%; Pred. No. 6.8e-206;
XX Matches 615; Conservative 9; Mismatches 38; Indels 185; Gaps 11;
Qy 1 MATKAVCVLKGDPVQGIINFEQKESNGPVKVGWSIKGLTEGLHGFHVHFGDNTAGCTS 60
Db 1 MATNPVCVLKGDPVQGIINFEQKESNGPVKVGWSIKGLTEGLHGFHVHFGDNTAGCTS 60
```

```
Qy 61 AGPHFNPLSRKHGPKDBERHVGDLGNVTADKGVADVSIEDSVISLSDHCIIIGRTLAV 120
Db 61 PGPHFNPLSRKHGPKDBERHVGDLGNVTADKGVADVSIEDSVISLSDHCIIIGRTLAV 120
Qy 121 HEKADDLGKGNBEESTKTGNAGSRLACGVIGIAQNLSNCNCISYPGHITGHR 173
Db 121 HEKADDLGKGNBEESTKTGNAGSRLACGVIGIR-----GTYYV-NHLPLRDMHNG 174
Qy 174 -----MAWKLSAA-----RTSGFVS----- 190
Db 175 RDLAVAVEPVVFSQMETKLITWADTAACGDIINGLPLVSARRGREILLGPADGMVSKGWR 234
Qy 191 LFAP-----GAKONETH-----VTG 205
Db 235 LLAPITAYAOOTRGLGCIITSITGRDKNVEGEVQIVSTAAQTFLATCIINGVCWTYH 294
Qy 206 GAAART-----TSGLT----- 216
Db 295 GAGTRTIASPKGPVIQMYTNVDQDLVGPASQGTSLTPCTCGSSDLYLVTRHADVIPVR 354
Qy 217 -----SLFSP-----GAS-----QNIQLIVDFIPVENLET 241
Db 355 RRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFRAAVCTRGVAKAVDFIPVENLET 414
Qy 242 TMRSPVFTDNSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLG 301
Db 415 TMRSPVFTDNSSPPVVPQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLG 474
Qy 302 FGAYMSKAHGIDPNIRGTGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDCHSTDA 361
Db 475 FGAYMSKAHGIDPNIRGTGRTITTTGSPITYSTYKFLADGGCGGAYDIIICDCHSTDA 534
Qy 362 TSLIGIGTVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 421
Db 535 TSLIGIGTVLDOAETAGARLVLATATPPGSVTVPHNIEVALSTTGEIPFYGKAIPLE 594
Qy 422 VIKGGRHLIFCHSKKCKDELAALVALGINAVAYRGLDYSVIPTSGDVVVVATDALMTG 481
Db 595 VIKGGRHLIFCHSKKCKDELAALVALGINAVAYRGLDYSVIPTSGDVVVVATDALMTG 654
Qy 482 YTGDFDSVIDNCTCTQTVDFSLDPTFTTITLTPQDAVSRTQRRGTRGKPGIYRFA 541
Db 655 YTGDFDSVIDNCTCTQTVDFSLDPTFTTITLTPQDAVSRTQRRGTRGKPGIYRFA 714
Qy 542 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWGVFT 601
Db 715 PGERPDMFSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWGVFT 774
Qy 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTP 661
Db 775 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPPSDQMWKCLIRLKPFLHGPTP 834
Qy 662 LLYRLGA 668
Db 835 LLYRLGA 841
```

Search completed: November 7, 2005, 20:10:27  
Job time : 136.969 secs

**This Page Blank (uspto)**

Result No.	Score	Query Match	Length	DB	ID	Description
1	2909.5	49.2	3011	1	GNWVC3	genome polyprotein
2	2850	48.2	3011	1	S40770	genome polyprotein
3	2821	47.7	3011	1	GNWVC5	genome polyprotein
4	2768.5	46.8	3010	1	A45573	genome polyprotein
5	2746	46.4	3010	1	GNWVCJ	genome polyprotein
6	2741.5	46.4	3010	1	GNWVTC	genome polyprotein
7	2737.5	46.3	3010	1	S18030	genome polyprotein
8	2725.5	46.1	3010	1	GNWVTW	genome polyprotein
9	2455.5	41.5	3014	1	JQ5620	genome polyprotein
10	2431	41.1	3033	1	TC1303	genome polyprotein
11	2428	41.1	3033	1	GNWVJ8	genome polyprotein
12	1980.5	33.5	386	2	S68016	ATPase/RNA helicase
13	1966.5	33.3	876	2	PC2219	polypeptide - hepa
14	1483.5	25.1	492	2	FS0326	polyprotein - hepa
15	1386.5	23.5	716	2	JQ1366	polyprotein - hepa
16	1067	18.0	216	2	S21337	genome polyprotein
17	1001	16.9	194	2	S06067	nonstructural prot
18	998	16.9	194	2	A54317	probable nonstruct
19	980	16.6	182	2	S32748	genome polyprotein
20	947	16.0	184	2	A61196	genome polyprotein
21	834.5	14.1	3005	2	T08841	polyprotein - dour
22	825.5	14.0	2970	2	T08839	polyprotein - marm
23	821	13.9	154	1	DSHUCZ	superoxide dismuta
24	807	13.7	179	4	T43640	superoxide dismuta
25	720	12.2	135	2	FS0327	polyprotein - hepa
26	719	12.2	135	2	FS0328	polyprotein - hepa
27	710	12.0	209	2	PC1306	genome polyprotein
28	693	11.7	135	2	FS0329	polyprotein - hepa
29	694.5	11.6	152	2	S36108	superoxide dismuta

Db	1186	TRGAKAVDPIPVENLETTMRSVFTDNSSPPVVFQVAHLHPTGSGSKTKVPAAYA	1245
Qy	284	AQGYKVLNPSVAATLGFAYMSKAHIDPNIRTVRTITTTGSPITYSTYKFLADGGC	343
Db	1246	AQGYKVLNPSVAATLGFAYMSKAHIDPNIRTVRTITTTGSPITYSTYKFLADGGC	1305
Qy	344	SGGAYDIIICDECHSDATSIIGTGTVDQAEATAGARLVVLATATPPGVTVPHPNIEEV	403
Db	1306	SGGAYDIIICDECHSDATSIIGTGTVDQAEATAGARLVVLATATPPGVTVPHPNIEEV	1365
Qy	404	ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV	463
Db	1366	ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKCDLAALVALGINAVAYRGLDVSV	1425
Qy	464	IPTSGLVVVATDALMTGYTGDPSVIDCNTCTVTQTVDFSLDPTFTIETITLPODAVSRT	523
Db	1426	IPTSGLVVVATDALMTGYTGDPSVIDCNTCTVTQTVDFSLDPTFTIETITLPODAVSRT	1485
Qy	524	QRRGRTGRGKPGIYRFVAPGERPSGMFDSVLCVCEYDAGCAWYELTPAETTVRLRAYMNT	583
Db	1486	QRRGRTGRGKPGIYRFVAPGERPSGMFDSVLCVCEYDAGCAWYELTPAETTVRLRAYMNT	1545
Qy	584	PGLPVQCDHLEFEGVFTGLTHIDAHPLSQTQKSGENLPYLVAQATVCARAQAPPSWD	643
Db	1546	PGLPVQCDHLEFEGVFTGLTHIDAHPLSQTQKSGENLPYLVAQATVCARAQAPPSWD	1605
Qy	644	QMKCLIRLKPHTLHGPTPLLYRLGAVONEITLTHPTVKYIMTQMSADLEVVTSS-----	696
Db	1606	QMKCLIRLKPHTLHGPTPLLYRLGAVONEITLTHPTVKYIMTQMSADLEVVTSSWLVVGG	1665
Qy	697	-----ACSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLA	735
Db	1666	-----ACSGKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLA	1725
Qy	736	EQPKQKALGL-----SRGKGAIPVDKEVLYQQVD-----EMECSSQAAPYIEQAQVIAHQ	786
Db	1726	EQPKQKALGLQATASQAE-VIAPAVQTNWQKLETFWAKHMMNFISIGIYLAGLSTLPG-1783	
Qy	787	FKEKVLGLIDNQVVVTP---DKELIYE-----AFDEMECASKAALI	826
Db	1784	-NPAIASLMAFTAAVTSPLTTSQTLLENILGGWVAQAAPGAATAPVAGLAGAAGISV	1842
Qy	827	EEGQRMALMLKSIQGLLG-----V-----IL	847
Db	1843	GLCKVLIDILAGYGAGVAGALVAFKIMSGEVSPFEDLVNLLPALISPGALVVGVCAAIL	1902
Qy	848	RRHVGPGEQAVQNMNRLIAFASRGNHVSPTHYPSS-----	882
Db	1903	RRHVGPGEQAVQNMNRLIAFASRGNHVSPTHYPSPESDAAAARVTAIISLTVTQLRLRHQ	1962
Qy	883	-----	882
Db	1963	WISSECTTPCGSGLRWLDIWDNI CEVLSDFTWLKAKLMPOLPGIPFVSCORGKGVWRVD	2022
Qy	883	-----	882
Db	2023	GIMHTRCHCAEITGHVQNGTMRIVGPRTCNMMSGTFFINAYTGTGCTPLPAPNYTFAL	2082
Qy	883	-----RSRRFA-----	888
Db	2083	WRVSABEYVEIRQVDPHYVTGTTDNLKCPQVPSPEFFTELDGVRHLHRFAPPCPKLLR	2142
Qy	889	-----QALPVWARPDY-----NPP-----	902
Db	2143	EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSSHITAEAGRRLARGSPSPVASSS	2202
Qy	903	-----LVET---WKK-----PDEYEPVWH	918
Db	2203	ASQLSAPSLKATCTANHSDSPAELIEANLLWRQEMGNNITRVESENKVVILDSFDPLVAE	2262
Qy	919	G-----RSSRRFAQLPVWARPDYNPPLVETWKKPDYEPVWHGRKTKRNTNR	966
Db	2263	EDREISVPAEILKRSRRFAQLPVWARPDYNPPLVETWKKPDYEPVWH-----	2312

Qy 967 RPQDVKPGCGQIVGRRGPPPIPKARR 992  
Db 2313 -----GCLPPPKSPVPVPPRK 2329

RESULT 2  
S40770

genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu  
Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: S40770; PC1285  
R;Okamoto, H.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S40770  
A;Accession: S40770  
A;Molecule type: genomic RNA  
A;Residues: 1-3011 <OKA>  
A;Cross-references: UNIPROT:Q03463; EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g22158  
R;Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tauda,  
Jpn. J. Exp. Med. 60, 167-177, 1990  
A;Title: The 5'-terminal sequence of the hepatitis C virus genome.  
A;Reference number: PC1284; MUID:91013116; PMID:2170712  
A;Accession: PC1285

A;Molecule type: genomic RNA  
A;Residues: 1-513 <OK2>  
A;Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512  
A;Experimental source: isolate HC-J1  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepacivirin #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>  
F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 48.2%; Score 2850; DB 1; Length 3011;  
Best Local Similarity 51.2%; Pred. No. 2.1e-169;  
Matches 608; Conservative 44; Mismatches 115; Indels 420; Gaps 18;

Qy	204	TGGAAARTTSGLTSLFSPG-ASQNIQLIIVDFIPVENLETTMRSVPFTDNSSPPVVPQSSEQ	262
Db	1165	SGGPLLCPAGHVGVIFRAAVCTRGVAKAVDFIPVESLETTMRSVPFTDNSSPPAVPQSSEQ	1224
Qy	263	VAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLGFAYMSKAHIDPNIRTVRT	322
Db	1225	VAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLGFAYMSKAHIDPNIRTVRT	1284
Qy	323	ITTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATSIIGTGTVDQAEATAGARLV	382
Db	1285	ITTGSPITYSTYKFLADGGCGGAYDIIICDECHSDATSIIGTGTVDQAEATAGARLV	1344
Qy	383	VLATATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLA	442
Db	1345	VLATATPPGSGVTVPHPNIEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCDLA	1404
Qy	443	AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDFSDVIDCNTCTVTQVDF	502
Db	1405	AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGFDFSDVIDCNTCTVTQVDF	1464
Qy	503	SLDPTFTIETITLPODAVSRTQRRGTGRKGCIVRFVAPGERPSPVSSVLCECYDAG	562
Db	1465	SLDPTFTIETITLPODAVSRTQRRGTGRKGCIVRFVAPGERPSPVSSVLCECYDAG	1524

563 CANYELTPAETTVRLRAYMNTPGLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENLP 622  
Db  
1525 CANYELTPAETTVRLRAYMNTPGLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENFP 1584  
Qy 623 YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTLLYRLGAVQNETTLTHPVTKY 682  
Db 1585 YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTLLYRLGAVQNETTLTHPVTKY 1644  
Qy 683 IMTCSADLEVVTS-----ACSGKPAIIPDREVLVRE 714  
Db 1645 IMTCSADLEVVTSVWLVGVLAAALAYCLSTGCVVIGRVLVLSGRPAIIPDREVLVRE 1704  
Qy 715 FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQYVD----- 765  
Db 1705 FDEMECSQHLPIYEQGMMLAEQFKQKALGLQTLQASRAE-VIAFTVQTNMQLKLEAFWAK 1763  
Qy 766 EMECSQAAPVIEQAQVIAHQFKEKVLGLINDQVVVTP---DKEIILYE----- 811  
Db 1764 HMNFIISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTLFNLILGGWVAQA 1821  
Qy 812 -----AFDEMECASKAALIEEGORMAELMSKIQGLLG----- 845  
Db 1822 APGAATFVSGSLAGAAVSGVGLGRVLVDILLAGYGAGVAGALVAFKIMSGELPSTEDLVN 1881  
Qy 846 -----ILRRHVGPEGAVQWNRLLIAFASRGNHVSPTHYVPS----- 882  
Db 1882 LLPAILSPGALVGVCAAILRRHVGPEGAVQWNRLLIAFASRGNHVSPTHYVPSDAA 1941  
Qy 883 ----- 882  
Db 1942 ARVTAIISLTVTLRLHQLWLSSESTPCSGSLRDINDWICEVLSDFKTLTKLMP 2001  
Qy 883 ----- 882  
Db 2002 HLGPIPVSCQHGKGVWRGDGIMHTRCHGAEITGHVKNGTMRIVGPKTCENMMWSTFP 2061  
Qy 883 ----- 882  
Db 2062 INAVTTGCTPLPAPNYTFALWRVSAEYVEIRRVDPHYVTGMTDNLKCPQCVPSPEF 2121  
Qy 883 -----RSRFA-----QALPVWARPY----- 899  
Db 2122 FTLDGVLRLHFAPECKPLLEEVSVFVGLHDYVPGSQLPCEPEPDVAVLTSMLTDPHSI 2181  
Qy 900 -----NBP-----LVET---WKK----- 909  
Db 2182 TAAAGRRLARGSPSEASSASQLSAPSLKATCTINHDSPDAELIEANLLWRQEMGNI 2241  
Qy 910 -----PDYEPVPHG-----RSSRFAQALPVWARPDPYNPLVETW 945  
Db 2242 TRVESENKVILDSFDPLVAEDEREISVPAEILKSRRTQALPIWARPDPYNPLIETW 2301  
Qy 946 KKPDPYEPVPHGKTKRNTNRRPDQVKFPGGQIVGRRGPPIPKARR 992  
Db 2302 KKPNYEPVPHGCPLP-----PPQ-----SPVPVPPRK 2329

RESULT 3  
GNMVCH  
N: genome polyprotein - hepatitis C virus (strain H)  
N: Contains: capsid protein C; envelope protein M; hepatitis C virus (strain H) of hepatitis C virus: C  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C: Species: hepatitis C virus  
A: Note: host Homo sapiens (man)  
C: Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C: Accession: A36814; A41546  
R: Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
submitted to GenBank, July 1992  
A: Description: Genomic structure of the human prototype strain H of hepatitis C virus: C  
A: Reference number: A36814  
A: Molecule type: genomic RNA  
A: Residues: 1-3011 <INC>

A: Cross-references: UNIPROT: P27958; GB: M67463; NID: G329737; PID: AAA45534.1; PID: G329738  
R: Inchauspe, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
A: Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
A: Reference number: A41546; MUID: 92052256; PMID: 1658800  
A: Contents: annotation  
A: Note: neither amino acid nor nucleotide sequence is given  
C: Superfamily: hepatitis C virus genome polyprotein  
C: Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F: 1-115/Product: capsid protein C #status predicted <CPC>  
F: 116-191/Product: envelope protein M #status predicted <EPM>  
F: 192-389/Product: major envelope protein E #status predicted <MEE>  
F: 390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F: 730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F: 1007-1615/Product: hepatitis C virus nucleocapsid protein NS3 #status predicted <NS3>  
F: 1230-1317/Product: nucleocapsid-binding motif A (P-loop)  
F: 1312-1317/Region: nucleocapsid-binding motif A (P-loop)  
F: 1316-1319/Region: DEXH motif  
F: 1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F: 1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F: 2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
F: 196,209,234,305,325,417,423,430,448,476,532,536,576,623,645,1213,1255,2041,2240,23  
Query Match 47.7%; Score 2821; DB 1; Length 3011;  
Best Local Similarity 51.3%; Pred. No. 1.4e-167;  
Matches 611; Conservative 41; Mismatches 110; Indels 430; Gaps 17;  
Qy 204 TGGAAARTTSLTSLFSPG-ASQNIQIIVDPYVENLETTMRSPVFTDSSPPVVPQSFQ 262  
Db 1165 SGGPLLCPTGHAVGLFRAAVCTRGVAKAVDIPVENLETTMRSPVFTDSSPPVVPQSFQ 1224  
Qy 263 VAHLHAPTSKSTKVPAAAYAAQYKVLVLPNSVAATLFGAYMSKAHGDIPNRTGVRT 322  
Db 1225 VAHLHAPTSKSTKVPAAAYAAQYKVLVLPNSVAATLFGAYMSKAHGDIPNRTGVRT 1284  
Qy 323 ITTCSPTSTYTGKFLADGCGCGGAYDIIICDECHSTDATSIIGTGLVDOAETAGARLV 382  
Db 1285 ITTCSPTSTYTGKFLADGCGCGGAYDIIICDECHSTDATSIIGTGLVDOAETAGARLV 1344  
Qy 383 VLATATPGSVTVPHNIEEVALSTTGTEIPFYKAIPLEVIKGRHLLFCHSKKCDLA 442  
Db 1345 VLATATPGSVTVSHNIEEVALSTTGTEIPFYKAIPLEVIKGRHLLFCHSKKCDLA 1404  
Qy 443 AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDPSVDCNTCTQTVD 502  
Db 1405 AKLVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTGYTGDPSVDCNTCTQTVD 1464  
Qy 503 SLDPFTTETITLQDASVRSRGRGRKPGYRFAVGPBPSGMPDSSVLCBCYDAG 562  
Db 1465 SLDPFTTETITLQDASVRSRGRGRKPGYRFAVGPBPSGMPDSSVLCBCYDAG 1524  
Qy 563 CANYELTPAETTVRLRAYMNTPGLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENLP 622  
Db 1525 CANYELTPAETTVRLRAYMNTPGLPVCDHLEFMEGVFTGLTHIDAHFLSQTQSGENFP 1584  
Qy 623 YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTLLYRLGAVQNETTLTHPVTKY 682  
Db 1585 YLVAQATVCARAQAPPSDQMKCLIRLKPDLHGPTLLYRLGAVQNETTLTHPVTKY 1644  
Qy 683 IMTCSADLEVVTS-----ACSGKPAIIPDREVLVRE 714  
Db 1645 IMTCSADLEVVTSVWLVGVLAAALAYCLSTGCVVIGRVLVLSGRPAIIPDREVLVRE 1704  
Qy 715 FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQYDEMESC 770  
Db 1705 FDEMECSQHLPIYEQGMMLAEQFKQKALGLQTLQASRAE-VITPAVQTNMQLKLEAFWAK 1762  
Qy 771 SOAAYPIBQAQVIAHQFKEKVLGLINDQVV---VTPDKKILYE----- 811  
Db 1763 KHMNFIISGIQYLAG-----LSTPGNPAIASLMAFTAAVTSPLTTSQTLFNLILGGWV 1816  
Qy 812 -----AFDEMECASKAALIEEGORMAELMSKIQGLLG----- 845

```
Db 1817 AAQLAAPGAATAFVAGLAGAALDSVGLKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876
QY 846 -----ILRRHVPGEGAVQWMNRLIAFASRGNHVSPTHYVP 881
Db 1877 EDLVNLLPALLSPGALAVGVFASILRRVRPGEGAVQWMNRLIAFASRGNHVSPTHYVP 1936
QY 882 S-----882
Db 1937 ESDDAAARVTAISSLVTQLLRRLHQWISSECTTPCGSGLRDIWDWICEVLSDFKTLWK 1996
QY 883 -----882
Db 1997 AKLMPQLPGIPFVSCQRYGVWRGDGIMHTRCHGAEITGHVKNGTMRIVGPRCKNMW 2056
QY 883 -----882
Db 2057 SGTFFINAYTGTCTPLDPAPNYKFAALWRVSAEYVEIRRVGDPHYVSGMTNMLKPCQI 2116
QY 883 -----RSRFFA-----QALPVWARPDPY-----899
Db 2117 PSPEFFTELDGVLHRFAPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLT 2176
QY 900 -----NPP-----LVET-----WKK- 909
Db 2177 DPSHITAAGRRRLARGSPSPSSASSASQLSAPSLKATCTANHDSFDAELIANLLWROE 2236
QY 910 -----PDYEPVPHG-----RSSRFFAQAALPVWARPDPNPP 940
Db 2237 MGNITRVSEKNVILDSFDPLVASEDEREVSVPAILRKRFRFAPALPVWARPDPNPL 2296
QY 941 LVETWKKPDYEPVPHGRTKRTNRRPDQVKFPGGQIVGRRGPIPKARR 992
Db 2297 LVETWKKPDYEPVPH-----GCPLEPPSPPPVPPPRK 2329

RESULT 4
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata,
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
A:Reference number: A45573; MUID:92295714; PMID:1318627
A:Accession: A45573
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: UNIPROT:Q00269; GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1;
A:Experimental source: HCV-JT
A>Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:790-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
F:1230-1237/Product: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 46.8%; Score 2768.5; DB 1; Length 3010;
Best Local Similarity 48.5%; Pred. No. 2.6e-164;
Matches 589; Conservative 56; Mismatches 136; Indels 434; Gaps 14;
```

```
QY 183 RTTSGEVSUFAPGAKQKONETHVTGGAARTTSLTSLFSPG-ASQNIQLIVDIPVENLET 241
Db 1144 RRGDGRGSLSPRPVSYLKGSSGPLLCPSHAVGIFRAAVCTRGVAKAVDIPVESMET 1203
QY 242 TWRSVPFTDNNSSPPVPQSFQVAHLHAPTGSCKSTKVPAAQAQYKVLVLNPSVAATLG 301
Db 1204 TWRSVPFTDNNSSPPVPQSFQVAHLHAPTGSCKSTKVPAAQAQYKVLVLNPSVAATLG 1263
QY 302 FGAYMSKAGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGGAYDIIICDECHSTD 361
Db 1264 FGAYMSKAGIDPNIRTGVRTITTTGSPITYSTYKFLADGGCGGGAYDIIICDECHSTD 1323
QY 362 TSIIGTGYLDOAETAGARLVVLATATATPGSVTVPHNPTEEVALSTGTBEIPYKAIPL 421
Db 1324 TSIIGTGYLDOAETAGARLVVLATATATPGSVTVPHNPTEEVALSTGTBEIPYKAIPL 1383
QY 422 VIKGBRHILFCHSKKKCDLAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTG 481
Db 1384 AIKGBRHILFCHSKKKCDLAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTG 1443
QY 482 YTGDPDSVIDCNTCTVTQTFDLSLDPFTTITLPODAVSRTOQRGRTGRGKPGIYRFVA 541
Db 1444 YTGDPDSVIDCNTCTVTQTFDLSLDPFTTITLPODAVSRTOQRGRTGRGKPGIYRFVT 1503
QY 542 PGERPSGMDSSVLCEDYDAGCAWYELTPTAETTVRLRAYMNTFGLPVQDHLFEFWEVFT 601
Db 1504 PGERPSGMDSSVLCEDYDAGCAWYELTPTAETTVRLRAYMNTFGLPVQDHLFEFWEVFT 1563
QY 602 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKLHPTLHGTP 661
Db 1564 GLTHIDAHFLSQTQSGENLPYLVAQATVCARAQAPPSWDMKCLIRLKLHPTLHGTP 1623
QY 662 LLYRLGAVONEITLTHPTKVTMTCMSADLEVVTSA-----697
Db 1624 LLYRLGAVONEITLTHPTKVTMTCMSADLEVVTSTWLVGGVLAALAAAYCLTGSWIV 1683
QY 698 ----CSGKPAIIPDREVLRYREFDEMECSQHLPIYEQGMMLAEQFKQKALGSRGKPAI 753
Db 1684 GRILLSGRPVAVPDREVLRYREFDEMECSQHLPIYEQGMMLAEQFKQKALGSLTATKQA 1743
QY 754 VPDKEVLYQOYDEME-----769
Db 1744 EAAAPVESRWRALAEFAWAKHMNFISGTYLAGUSTLPGNPAIASLMAFTASITPLTT 1803
QY 770 -----CSQAAPYIEQAQVIAHQFKEKVLGLIDNDQVVVTPDKEIL-----809
Db 1804 QNTLLFNILGGWVAQALAPPSAFAVGAGIAGAAIGSIGLKVLV----DILAGYGAV 1859
QY 810 ----YEAFDEMECSAKA-----ALIEEGORMAEMLSKTKQGLLGLIRRHVGPGEAV 858
Db 1860 AGALVAFKVMGSEAPSAEDLVNLLPALLSPGALVGVW-----CAAILRRHVGPGEAV 1913
QY 859 QMNRLLIAFASRGNHVSPTHYVP-----SRSRFAQALPV-----WARDYNPPLV 904
Db 1914 QMNRLLIAFASRGNHVSPTHYVPESDAAARVTOILSSLTITOLLKRLHQMINECDSTPCS 1973
QY 905 ETWKK-----909
Db 1974 GSWLKDVNDWICTVLTDFKTLWLSQKLLPKLPGVFPFCQRYGVWRGDGIMQTTCPGA 2033
QY 910 -----PDY-----912
Db 2034 QITGHVKNMIRIVGPKTCNTWHGTFFINAVTTGPTSPAPNYSRALWRVAABEYVEI 2093
QY 913 -----912
Db 2094 TRVGDFHYVTGMTDNVKPCQVPAPEFFTELDGVLRLHYAPACRPLLRDVTTFQVLNQ 2153
QY 913 -----BPPVV-----917
Db 2154 YLVGSQLPCEPEPDVAVLTSMLTDPSSHITAEAKRRLARGSPSPSSASSASQLSAPSLKA 2213
QY 918 -----HG-----919
```



Db	2214	TCCTHSDPADLIEANLLMRQMGNIITRVESENKVVILDSFDPLRAEBEDREVSAAE	2273
Qy	920	--RSSRFAQALPVYARVDYNNPLVETWKKDDYPPVHVGHGKTKRNTNRRPQDVKFFGG	977
Db	2274	ILRSKFPFPPALPIWARPVDYNNPLLESKSDYVPPAVHGCPLPPTT-----	2320
Qy	978	QIVGRRGPPPIPKARR	992
Db	2321	-----GPIPPPRK	2329
RESULT 5			
GNWVCJ			
Genome polyprotein - hepatitis C virus (strain J)			
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004			
C:Accession: A39253; PS0086			
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto			
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990			
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients			
A:Reference number: A39253; MUID:91088550; PMID:2175903			
A:Accession: A39253			
A:Molecule type: genomic RNA			
A:Residues: 1-3010 <KAT>			
A:Cross-references: UNIPROT:P26662; GB:D90208; NID:G221610; PIDN:BAAL14233.1; PID:G221611			
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.			
Proc. Jpn. Acad. 65B, 219-223, 1989			
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari			
A:Reference number: PS0085			
A:Accession: PS0086			
A:Molecule type: genomic RNA			
A:Residues: 2650-2707 <KAT>			
A:Experimental source: Japanese isolate			
C:Comment: The cleavage sites of this polyprotein have not been determined.			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin			
F:2-115/Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <EPM>			
F:192-389/Product: major envelope protein E #status predicted <MEE>			
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>			
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>			
F:1230-1237/Region: nucleotide-binding motif A (P-loop)			
F:1312-1317/Region: nucleotide-binding motif B			
F:1316-1319/Region: DEXH motif			
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>			
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>			
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 2			
Query Match 46.4%; Score 2746; DB 1; Length 3010;			
Best Local Similarity 48.2%; Pred. No. 6.7e-163;			
Matches 579; Conservative 58; Mismatches 116; Indels 448; Gaps 14;			
Qy	204	TGCAARTSGLTSLFSPG-ASQNIQLIVDPIPVENLETMTESPVFTDNSSPPVVFQSFQ	262
Db	1165	SGGPLLCPGHHVGFRAAVCTRGVAKAVDFIPVSMETMTESPVFTDNSSPPVVFQSFQ	1224
Qy	263	VAHLHAPTSGSKSTKVPAAAYAGQYKVLNPSVAATLFGAYMSKAHGIDPNIRTVRT	322
Db	1225	VAHLHAPTSGSKSTKVPAAAYAGQYKVLNPSVAATLFGAYMSKAHGIEPNIRTVRT	1284
Qy	323	ITTGSPITVSTYKGLADGCGSGGAYDIIICDECHSDATSLIGTGLVLDQAEAGARLV	382
Db	1285	ITTGSPITVSTYKGLADGCGSGGAYDIIICDECHSDSTDTILGIGTGLVLDQAEAGARLV	1344
Qy	383	VLATATPPGSVTVPHNIEEVALSTGEIPFYKAIPLVIGKGRHLIFCHSKKKDELA	442
Db	1345	VLATATPPGSIIVPHNIEEVALSNTGEIPFYKAIPIEAIKGRHLIFCHSKKKDELA	1404

RESULT 6

GNWVCJ

Genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural

Qy	443	AKLVALGINAVYVRGLDVSVIPTSGDVVVVATDALMTGYTGDSDSVIDCNCVTQTQVDF	502
Db	1405	AKLTLGLNNAVYVRGLDVSVIPTSGDVVVVATDALMTGYTGDSDSVIDCNCVTQTQVDF	1464
Qy	503	SLDPTFTIETITLQDQAVSRTRGRGKPGIYRFVAPGGRPSGMFSSSVLCECYDAG	562
Db	1465	SLDPTFTIETITLQDQAVSRTRGRGKPGIYRFVAPGGRPSGMFSSSVLCECYDAG	1524
Qy	563	CWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGLTHDAHFSLQTKSGENLP	622
Db	1525	CWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGLTHDAHFSLQTKSGENLP	1584
Qy	623	YLVAQATVCARACAPPSMDQMKCLIRLKTPLHPTPLLYRLGAVONEITLTHPTVKY	682
Db	1585	YLVAQATVCARACAPPSMDQMKCLIRLKTPLHPTPLLYRLGAVONEITLTHPTVKY	1644
Qy	683	IMTQMSADLEVVVTS-----CSGKPAIIPDREVLVRE	714
Db	1645	IMTQMSADLEVVVTS-----CSGKPAIIPDREVLVRE	1704
Qy	715	PDMEBESQHLPIYIEQGMMLAEQKQKALGSLRGKPAIVDPKEVLVYQYDEMEECSAA	774
Db	1705	PDMEBESQHLPIYIEQGMMLAEQKQKALGSLRGKPAIVDPKEVLVYQYDEMEECSAA	1747
Qy	775	PYIEQAQ-----VIAHQFKEKVLGLIDNDQVV-----VTPDKKEI	808
Db	1748	PVSESKVRALVEFWAKMMNFISIGIYLAGUSTLPGNPAIASLMAFTASITSPLTQNTL	1807
Qy	809	LY-----EAFDEMEECASKAALIEGORMAEMLKSKIQGLG-----	845
Db	1808	LFNILGGWVAQAAPPSSAASAFVAGIAGAVSGIIGKVLVDILAGYAGVAGALVAPK	1867
Qy	846	-----ILRRHVGPGEQAVQMMRLIAPASRGN	872
Db	1868	VMSGEMPSTDLNLLPAILSPGALVGVCAAILRRHVGPGEQAVQMMRLIAPASRGN	1927
Qy	873	HVSTHYVP-----SRSRFAQALPV-----WAPDYNPLVETWKK-----	909
Db	1928	HVSTHYVPESDAAARVTQILSSLTITQLKRLHQWINEDCSTPCSGSNLKDQWWDWICTV	1987
Qy	910	-----PDY-----	912
Db	2048	GPKTCSTNTHGCTFPINAYTTGCTPSAPNYSRALWRVAEBEYVEVTRVGFHYVTGMTT	2107
Qy	913	-----BPP	915
Db	2108	DNVKKPCQVPAPEFFTEVDGVLRLHRYAPVCKPLLEEVVQVGLNQYLVGSOLPCEPED	2167
Qy	916	VV-----HG-----	919
Db	2168	VAVLTSMLTDPSHITAETAKRLARGSPPSLASSASQLSAPSLKATCTTHHDSPADLI	2227
Qy	920	-----RSSRFAQALPV	931
Db	2228	EANLLWRQMGNIITRVESENKVVILDSFDPIRAVEDEREISVPAEILKPKRFPFALPI	2287
Qy	932	WAPDYNPLVETWKKPDYEPVHVGRKTKRNTNRRPQDVKPPGGOIVVGRGPPIPKAR	991
Db	2288	WAPDYNPLVETWKKPDYEPVHVGRKTKRNTNRRPQDVKPPGGOIVVGRGPPIPKAR	2328
Qy	992	R	992
Db	2329	R	2329

RESULT 6

GNWVCJ

Genome polyprotein - hepatitis C virus

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: A38465  
R;Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;  
J. Virol. 65, 1105-1113, 1991  
A;Title: Structure and organization of the hepatitis C virus genome isolated from human  
A;Reference number: A38465; MUID:91140698; PMID:1847440  
A;Accession: A38465  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <TAK>  
A;Cross-references: UNIPROT:P26663; EMBL:M58335; NID:g329770; PIDN:AAA72945.1; PID:g329770  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F;2-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEE>  
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1007-1615/Product: hepatitis C virus genome polyprotein NS3 #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1316-1319/Region: DEXH motif  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>  
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 46.4%; Score 2741.5; DB 1; Length 3010;  
Best Local Similarity 48.9%; Pred. No. 1.3e-162;  
Matches 577; Conservative 56; Mismatches 100; Indels 447; Gaps 14;

QY 224 SONIQLVDFIPVENLETTWRSPVFTDNSSPVVQSQFVAHLHAPTKSGSKTKVPAAYA 283  
DB 1186 TRGAKAVDFPVVSMETWRSPVFTDNSSPVVQSQFVAHLHAPTKSGSKTKVPAAYA 1245  
QY 284 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVRTITTSPTITYSTYGRFLADGGC 343  
DB 1246 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVRTITTSPTITYSTYGRFLADGGC 1305  
QY 344 SGAYDIIICDECHSTDATILGIGTVLDQAGTARLVVLATATPGSVTVVPHNIEEV 403  
DB 1306 SGAYDIIICDECHSTDATILGIGTVLDQAGTARLVVLATATPGSVTVVPHNIEEV 1365  
QY 404 ALSTTGTIPYGYKAIPLEVIKGGRHLLFCHSKKCDLAALVALGINAVAYRGLDVSV 463  
DB 1366 ALSTTGTIPYGYKAIPLEVIKGGRHLLFCHSKKCDLAALVALGINAVAYRGLDVSV 1425  
QY 464 IPTSGDVVVVATDALMTGYTGDFSDVIDCNCVTQTVDVDFSLDPTFTTITLTPQDAVSR 523  
DB 1426 IPTSGDVVVVATDALMTGYTGDFSDVIDCNCVTQTVDVDFSLDPTFTTITLTPQDAVSR 1485  
QY 524 QRRGTRGRGPGIYRFVAPGRSGMDFSSVLCBCEYDAGCAWYELTTPAETTVRLAYMNT 583  
DB 1486 QRRGTRGRGPGIYRFVAPGRSGMDFSSVLCBCEYDAGCAWYELTTPAETTVRLAYMNT 1545  
QY 584 PGLPVCDHLEFEGVETGLTHIDAHFLSOTKSGENLPYLVAQVATVCARAQAPPPSWD 643  
DB 1546 PGLPVCDHLEFEGVETGLTHIDAHFLSOTKSGENLPYLVAQVATVCARAQAPPPSWD 1605  
QY 644 QMWKCLIRLPTLHGPTPLLYRLGAVQNEITLHPVTVKYIMTCSADLEVVTSIA----- 697  
DB 1606 QMWKCLIRLPTLHGPTPLLYRLGAVQNEITLHPVTVKYIMTCSADLEVVTSIA----- 1665  
QY 698 -----CSGKPAIIPDREVLVREDFEMECESQHLPIYEQGMMLA 735  
DB 1666 VLAALAAAYCLTGSWIVGRILISGRPAIVPDRELLYQEFDEMEECAHSLFYIEQGMMLA 1725  
QY 736 EQFKQKALGSRGKPAIVDPKVELYQYDEMEECSOAPVIEQAQVIAHQFKEK-VLGL 794  
DB 1726 EQFKQKALGSLQ-----TATKQEAAPVSVESKWALETFWAKHWNF 1768  
QY 795 IDNDQVV-----VTPDKELIY-----EA 812

Db 1769 ISGIOYLAGLSTLPGNPAIASLMAFTASITSTLTTQSTLLFNILGWAQAPPSAASA 1828  
QY 813 FDEMEECASKAALIEBGGQMAEMLKSKIOGLLG----- 845  
Db 1829 FVAGIAGAAVSGSIGLUGKLVLDILAGYGAGVAGALVAFKVMGSEMPSTEDLVNLLPAILS 1888  
QY 846 -----ILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPS----- 882  
Db 1889 PGALVVGVCVCAAILRRHVGPGEGAVQMMNRLIAFASRGNHVSPTHYVPSDAAARVTQIL 1948  
QY 883 ----- 882  
Db 1949 SSLITQTLKRLHQMINECDSTPCSGSLRDVMDWICTVLDTFKTLQSKLLPOLPGVPF 2008  
QY 883 ----- 882  
Db 2009 FSCQGYKGVWRGDGIMQITPCGAQITGHVKNGSMRIVGPKTCSTWHTGTFPINAYTTG 2068  
QY 883 ----- 882  
Db 2069 PCTPSPAPNYSRALWRVAEEVETRVGDFHYVTGTTDNVKKPCQVPAPEFSEVDGV 2128  
QY 883 RSRPFAQA-----LPVWARPDY----- 899  
Db 2129 RLHRYAPACRPLLRREBVTQVGLNQYLVGSQLPCEPEPDVAVLTSLMTPDPSHITAETAKR 2188  
QY 900 -----NPP-----LVET-----WKK----- 909  
Db 2189 RLARGSPPLSSASSASQLSAPLSKATCTTHVHSPDADLEIANLLWQEMGNTRVSESN 2248  
QY 910 -----PDYEP-----PVVHGRSSRRFAQALPVWARPDPYNNPPLVETWKKPDYEP 952  
Db 2249 KVVVLDSPDLRAEEDEREVSVAETLRKSKFPAAMPVIAWARPDPYNNPPLLESWKDPDYVP 2308  
QY 953 PVVHGRKTRNTNRRPDQVKFGGQIVGRRGPPPIPKARR 992  
Db 2309 PVVH-----GCPLPPIKAPPPIPPRR 2329

RESULT 7  
S18030  
genome polyprotein - hepatitis C virus (isolate JK1)  
N;Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural  
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
A;Variety: isolate JK1  
C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C;Accession: S18030; S33570; A48332; S18029  
R;Honda, M.; Kaneko, S.; Masahashi, U.; Kobayashi, K.; Murakami, S.  
submitted to the EMBL Data Library, September 1991  
A;Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient  
A;Reference number: S18028  
A;Accession: S18030  
A;Molecule type: genomic RNA  
A;Residues: 1-3010 <HON>  
A;Cross-references: UNIPROT:Q68949; EMBL:X61596; NID:g59478; PIDN:CAA43793.1; PID:g59479  
A;Experimental source: isolate JK1 from an individual  
R;Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.  
Arch. Virol. 128, 163-169, 1993  
A;Title: Sequence analysis of putative structural regions of hepatitis C virus isolated from  
A;Reference number: A48332; MUID:93119270; PMID:8380322  
A;Accession: S33570  
A;Molecule type: genomic RNA  
A;Residues: 1-547, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>  
A;Cross-references: EMBL:X61591  
A;Note: this sequence is inconsistent with the nucleotide translation  
as Trp, and TTC for residue 771 as Ser  
A;Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
F;2-115/Product: capsid protein C #status predicted <CPC>



QY 263 VAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDPNIRTVGT 322  
Db 1225 VAHLHAPTCGSKSTKVPAAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDPNIRTVGT 1284  
QY 323 ITTGSPITYTYGKFLADGGCGGAYDIIICDBCHSTDATSIIGTIGTVDLQDAETAGARLV 382  
Db 1285 ITTGAPITYTYGKFLADGGCGGAYDIIICDBCHSTDSITLIGITGTVLDQDAETAGARLV 1344  
QY 383 VLATATPPGSVTVPHPNIEEVALSTTGEIIPFYGKAIPLEVIKGRHLIFCHSKKKKDELA 442  
Db 1345 VLATATPPGSVTVPHPNIEEIALSNTGEIPFYGKAIPETIKGRHLIFCHSKKKKDELA 1404  
QY 443 AKLVALGINAVAYRGLDVSVPTSGDVVVVATDALMTGTYGDPFDSVIDCNTCVTQTVDF 502  
Db 1405 AKLSALGIIHAYAYRGLDVSVTPASGNVVVATDALMTGTYGDPFDSVIDCNTCVTQTVDF 1464  
QY 503 SLDPFTTETITLPODAVSRTRGRGTGKPGIYRFVAPGERPSQMFSSVLCYCYDAG 562  
Db 1465 SLDPFTTETITWPDQAVSRQRGRTSRGRGIYRFVTPGERPSQMFSSVLCYCYDAG 1524  
QY 563 CAWYELTPAETTVRLRAYMNTPLGVCQDHLFEWGFVTLGTHIDAHFLSQTQKSGENLP 622  
Db 1525 CAWYELTPAETSVRLRAYLNTPLGVCQDHLFEWESVFTGLTHIDAHFLSQTQKAGDNFP 1584  
QY 623 YLVAYQATVCARAQAPPSQDQWKCLIRLKPTLHGPTLLYRLGAVQNEITLTHPTVKY 682  
Db 1585 YLVAYQATVCARAQAPPSQDQWKCLIRLKPTLHGPTLLYRLGAVQNEVTLTHPTIKY 1644  
QY 683 INTCMSADLEWTSIA-----CSGKPAIIPREVLYRE 714  
Db 1645 IMACMSADLEWTSITWVLGGVLAALAAAYCLTTGVSIVVIGRIILSKPAPVPREVLYQE 1704  
QY 715 FDEMECSOHLPIYEQGMMLAEQFKQALGLSRGKPAIVDPKEVLYQQYDEMECSQAA 774  
Db 1705 FDEMEECASHLPIYEQGMQALAEQFKQALGLQ-----TATKQAEAAA 1747  
QY 775 PYIEQAQ-----VIAHQFEKVLGLINDQV-----VTPDKEI 808  
Db 1748 PVYESKRWTLFAPWANDMNFISGIQYLAGSLTLPGNPAIASLMAFTASITSLTQSTL 1807  
QY 809 LY-----EAFDEMECASKAALIEEGORMAMLKSKTQGLG-----845  
Db 1808 LFNILGGVNAQAAPPGAASAFVAGIAGNAGVCSIGLGKVLVDVNAVAGYAGVAGALVAFK 1867  
QY 846 -----ILRRHVGPGEQVQMMNRLIAFASRGN 872  
Db 1868 VMSGEMPSTEDLVNLLPAILSPCALVGVVCAAILRRHVDPGEGAVQMMNRLIAFASRGN 1927  
QY 873 HVSPTHYVPS-----882  
Db 1928 HVSPTHYVPESDAAARVTOILSGLTITQLLRLHQWINEDECSTPCSGMRLRDVMDWICTV 1987  
QY 883 -----882  
Db 1988 LADFKTWLOSLPLRPLPGVPFFSCQRYGKVRWGDGIMQTTCPGGAQLTGHVKNXGSMRIW 2047  
QY 883 -----882  
Db 2048 GPKTCSNTWHGTFFPINAYTTGCTPSPAPNYSRALMRVAEEVVEVRVGDHFVYTGMTT 2107  
QY 883 -----RSRRFAQA-----LPVWARPD 898  
Db 2108 DNVKPCQVPAPBFFTEVDGVRHLRYAPACKPLLRBEVSFQVGLNQYVWGSQLPCEPEPD 2167  
QY 899 -----YNPP-----LV 904  
Db 2168 VAVLTSMLTDPSSHITAETAKRRLARGSPPLASSASQLSALSAACTTRHTHPPADLI 2227  
QY 905 ET---WKK-----PDYEP-----PVHGSSRRFAQALPV 931  
Db 2228 EANLWRQMGGNITRVESENKVILDSFDFLRAEEDEREVSVPABILRKSRRKFPFAPALPV 2287

QY 932 WARPDYNPLVETWKKPDYEPVPHVGRKTKRNTNRRPQDVKPPGGQIQVGRGPPPIPKAR 991  
Db 2288 WARPDYNPLLEPFWKDPDYVPPVPH-----CCPLPPVKAPPIPPR 2328  
QY 992 R 992  
Db 2329 R 2329  
RESULT 9  
JC5620  
genome polyprotein - hepatitis C virus (isolate EUH1480)  
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
C:Species: hepatitis C virus  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 09-Jul-2004  
C:Accession: JC5620  
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
A:Reference number: JC5620; MUID:97366593; PMID:9223423  
A:Accession: JC5620  
A:Molecule type: mRNA  
A:Residues: 1-3014 <CHA>  
A:Cross-references: UNIPROT:O39928; GB:Y13184  
A:Experimental source: genotype 5a, which predominates in South Africa  
A:Note: the translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: hepatitis C virus genome polyprotein  
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serine  
F:2-115/Product: capsid protein C #status predicted <CPC>  
F:116-191/Product: envelope protein M #status predicted <EPM>  
F:192-389/Product: major envelope protein E #status predicted <MEB>  
F:384-408/Region: hypervariable #status predicted  
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
F:1008-1616/Product: hepatitis C virus #status predicted <NS3>  
F:1231-1238/Region: nucleotide-binding motif A (P-loop)  
F:1313-1318/Region: nucleotide-binding motif B  
F:1317-1320/Region: DEH motif  
F:1617-1863/Product: nonstructural protein NS4a #status predicted <NS4>  
F:1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>  
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  
F:2210-2249/Region: interferon sensitivity determining #status predicted  
Query Match 41.5%; Score 2455.5; DB 1; Length 3014;  
Best Local Similarity 42.9%; Pred. No. 18-144;  
Matches 531; Conservative 85; Mismatches 172; Indels 450; Gaps 17;  
QY 179 GSAARTTSGFVSLF-----APGAKQNETHV-----TCGAAARTTSGLT 216  
Db 1119 GSLTRCTCGSADLYLVTRHADVIPARRGDTTRASLLSPRISVLKSGSGPIWCPSGHV 1178  
QY 217 SLFSPG-ASONTQLIVDFIPVENLETMRSPVFTNNSPPVPPQSPQVAHLHAPTSGSKS 275  
Db 1179 GVFRAAVCTRGVAKALEFVVENLETMRSPVFTNNSPPVPPQSPQVAHLHAPTSGSKS 1238  
QY 276 TKVPAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDPNIRTVGTITGSPITYSTVG 335  
Db 1239 TKVPAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDPNIRTVGTITGSPITYSTVG 1298  
QY 336 KFLADGGCGGAYDIIICDBCHSTDATSIIGTIGTVDLQDAETAGARLVLATATPPGSVTV 395  
Db 1299 KFPADGGCGGAYDIIICDBCHSQDATTILGIGTVDLQDAETAGARLVLATATPPGSVTT 1358  
QY 396 PHPNIEEVALSTTGEIIPFYGKAIPLEVIKGRHLIFCHSKKKKDELAALVALGINAVAY 455  
Db 1359 PHPNIEEVALPSEGEIIPFYGRAIPLVLIKGRHLIFANOKKAKETAKONKPGREKAVY 1418  
QY 456 YRGLDVSVPTSGDVVVVATDALMTGTYGDPFDSVIDCNTCVTQTVDFSLDPTFTITL 515  
Db 1419 YRGLDVAVIPATGDDVVVVCSTDALMTGTYGDPFDSVIDCNTSAVQTQTVDFSLDPTFTITTV 1478  
QY 516 PQDAVSRTRGRGTGKPGIYRFVAPGERPSQMFSSVLCYCYDAGCAWYELTPAETTV 575



Db	1839	VGAAGSIGLKVLDILACYGAGISGALVAFKIMSGKPSMEDVNVNLLPGLISPGALVV	1898
QY	845	-----GILRRHVGPGEAGVQWNRLIAFASRGHNHVSPTHV--PSRERRFAQAL-----	891
Db	1899	GVICAALLRHVGPGEAGVQWNRLIAFASRGHNHVAPTHVTHVTSADASQRTVQLLSLIT	1958
QY	892	-----	891
Db	1959	SLLRRLHNWITDCPIPCSGSLRDLVMDVCTILTDFKNLTSKLPKMPGLPFISQKG	2018
QY	892	--PVNA-----	895
Db	2019	YKQVWAGTGMTTRCPGCGANISGNVRLGSMRITGPCTKMNWQGTFFPINCYTEGCQVPKP	2078
QY	896	RPDYNPLVETWK-----	908
Db	2079	APNFK---IAIWRVAASEYAEVTHQGSYHYITGLTTLNKLVPQLSPPEFSSWVDGVQIH	2135
QY	909	-----KP-----DYPE-----	914
Db	2136	RFAPIPKPPRDEVSFCVGLNSFVGSQPCDPEPDTVLTSMLTDPSHITAETAARRLA	2195
QY	915	-----	914
Db	2196	RGSPPEASASQLSAPSIRATCTTHGKAYDMDVDANLFGMGDVTRIESESKVYVLD	2255
QY	915	--PVMHGRS-----SRRFAQALPVWARPDPYNPLVETWKKPDYEPVHVH--	957
Db	2256	LDPWVERSDLESISPEYMLPKRPPALPAWARPDPYNPLVESWKRPDYQATVAGCA	2315
QY	958	-----RKTNRNRRPDQVKF-----PGGQIV-----GRR	983
Db	2316	LPPEKXTPTPPRRRTVGLSESIADALQOLAISFGQPPPSGDSGLSTGADAADSGSR	2375
QY	984	GPPIKARRPGEWTWQGPVWPLYGNKDRRTSKGKPGYWPWRKTKRNTNRRPDQVK	1043
Db	2376	TPDELALSTGTSISSMP-----PLE-----GEPGDP-----DLEPEQVE	2410
QY	1044	F---PGGQIVGRRG	1055
Db	2411	LQPPPGGVVTPGSG	2425
RESULT 11			
GNMVJ8			
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain HC-J8)			
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5			
C:Species: hepatitis C virus			
C>Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 09-Jul-2004			
C:Accession: A40250; PQ0397; PQ0559			
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;			
Virology 188, 331-341, 1992			
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep			
A:Reference number: A40250; MUID:92230232; PMID:1314459			
A:Accession: A40250			
A:Molecule type: Genomic RNA			
A:Residues: 1-3033 <OK>			
A:Cross-references: UNIPROT:P26661; GB:D10988; GB:D01221; NID:g221608; PIDN:BAA01761.1;			
R:Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.H.			
J. Gen. Virol. 73, 1131-1141, 1992			
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e			
A:Reference number: PQ0393; MUID:92268871; PMID:1316939			
A:Accession: PQ0397			
A:Molecule type: Genomic RNA			
A:Residues: 2678-2754 <CHA>			
A:Cross-references: DDBJ:D10134			
A:Experimental source: isolate E-b12			
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnd			
Biochem. Biophys. Res. Commun. 181, 279-285, 1991			
A:Title: Distribution of plural HCV types in Japan.			
A:Reference number: PQ0554; MUID:92068204; PMID:1720309			
A:Accession: PQ0559			

A:Molecule type: mRNA			
A:Residues: 2678-2729 <KAT>			
A:Cross-references: GB:D10562; GB:D90518; NID:g221523; PIDN:BAA01418.1; PID:g221524			
C:Superfamily: hepatitis C virus genome polyprotein			
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural			
F:1-115/Product: capsid protein C #status predicted <CPC>			
F:116-191/Product: envelope protein M #status predicted <EPM>			
F:192-389/Product: major envelope protein NS1 #status predicted <NS1>			
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>			
F:1011-1619/Product: hepatitisin #status predicted <NS3>			
F:1234-1241/Region: nucleotide-binding motif A (P-loop)			
F:1316-1321/Region: nucleotide-binding motif B			
F:1320-1323/Region: DEXH motif			
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>			
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>			
F:2018-2033/Product: nonstructural protein NS5 #status predicted <NS5>			
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,231;			
Query Match 41.1%; Score 2428; DB 1; Length 3033;			
Best Local Similarity 41.9%; Pred. No. 5.3e-143;			
Matches 543; Conservative 54; Mismatches 128; Indels 570; Gaps 18;			
QY	204	TGGAARTTSGTSLFSPG-ASQNIQLIVDFIPVENLETMTMRSPVTDNSSPPVPPQSFQ	262
Db	1169	SGGPVLCSRGHAVGLFRAAFCARGVAKSIDFIPVESLDVATRTPSFSDNSTPPAVPQSYQ	1228
QY	263	VAHLHAPTCSGSKTKVPAAYAAQYKVLVNPVAATLFGAYMKAHGIDPNIRTVRT	322
Db	1229	VGLHAPTCSGSKTKVPAAYASQGYKVLVNPVAATLFGAYMKAHGIDPNIRTVRT	1288
QY	323	ITGSPITVSTYKFLADGGCGSGAYDIIICDEHSTDATSILGICTVLDDAETAGARLV	382
Db	1289	VITGDSITVSTYKFLADGGCGAAGAYDIIICDEHSDVATTILGICTVLDDAETAGARLV	1348
QY	383	VLATATPPGSVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA	442
Db	1349	VLATATPPGTVTPHNSNIEEVALGHEGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA	1408
QY	443	AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGTGDFDSVDCNTCTVTQVDF	502
Db	1409	AALRGMGVNAVAYRGLDVSIVPTQGDVVVVVATDALMTGTGDFDSVDCNVAVSQIVDF	1468
QY	503	SLDPTFTIETITLPQDAVSRTORRGTRGKPGIYRFVAPGERPSPGMFSSVLCECYDAG	562
Db	1469	SLDPTFTIITQVPQDAVSRSQRRGTRGRLGVYRVSSGERPSPGMFSDVLCICYDAG	1528
QY	563	CAWYELTPAETTVRLRAYNWTGRLPVCQDHLBFWEVFTGLTHIDAHFLSQTKSGENLP	622
Db	1529	AAWYELTPAETTVRLRAYNWTGRLPVCQDHLBFWEAVFTGLTHIDAHFLSQTKSGENFA	1588
QY	623	YLVAQATVCARQAAPPSPSDQMWKLIQLKPTLHGPTLLYRLGAVQNEITLTHPVTKY	682
Db	1589	YLTAYQATVCARAKAPPPSDVWVWKKLTKLTKPTLTGPTLLYRLGAVTNEVTLTHPVTKY	1648
QY	683	IMTMSADLEWVTSACSGKPAIIPDREVLRYRFDEMECSQHLPIYEQGMMLAEQFKQA	742
Db	1649	IATCMQADLEIMTSS-----VAAYCLATGCIS-----IIGRLHLNDRVV	1672
QY	743	LGLSRGGKPAIVDPKEVLYQQYVDEMECSQAAPYIEQAQVIAHQFKKVKGLLID-NDQVV	801
Db	1673	-----VAAYCLATGCIS-----IIGRLHLNDRVV	1697
QY	802	VTPDKELYEAFDEMECSKAALTEEGORMEMLKSKIQGLL-----	844
Db	1698	VAPDKELYEAFDEMECSKAALTEEGORMEMLKSKIQGLLQATROAQDIQPIAQSS	1757
QY	845	-----	844
Db	1758	WPKLEQFWAKHMWNFISGTYLAGLSTLPGNPVAVASMMAFSAALTSPLTSTILLINMG	1817
QY	845	-----	844

Db 1818 GWLASQIAPPAGATGFVVSGLVGAAGVSGIGLKILVDVLVLAGYAGAGISGALVAFKINSGEK 1877  
Qy 845 -----GILRRHYGPGEGAGVQWMNRLIAFASRGNHVSPTH 878  
Db 1878 PTVEDVVNLLPAILSPGALVGVICAAILRRHYGQEGAGVQWMNRLIAFASRGNHVAPTH 1937  
Qy 879 YV--PSRRRPAQALP----- 892  
Db 1938 YVVEDASORVTOVLSLTTITSLRLRLHAWITEDCPVPCSGSLQDIWDWCSILTDFKN 1997  
Qy 893 -----VWA----- 895  
Db 1998 WLSKLLPKMPGIPFISQKGYKGVWAGTGMVTRPCPGANISGHVRKMTKITGPKTCL 2057  
Qy 896 -----RPDYNPLVET--WK----- 908  
Db 2058 NLWQGTPIPCYTEGPCVKP-----PPNYKTAIRWAASEYVEVTQHGSFSYVTLGTSNLT 2114  
Qy 909 -----KPDYE-- 913  
Db 2115 KVPQVPAPEFFSWVDGVQIHRFAPVPGPFPRDEVTFTVGLNSFVVGSQLPCDPEDPTEV 2174  
Qy 914 -----PPVHGRSS----- 922  
Db 2175 LASMLTDPSHITABAAARRLARGSPSQASSASQLSAPSLKATCTTHKTAAYDCMVDAN 2234  
Qy 923 -----RRFAQALPVWARPDPYN 938  
Db 2235 LPWGGDVTRIESDKVIVLSDLSMTVEDREPSVPSEVYLKRRKPPALPPWARPDPYN 2294  
Qy 939 PPLVETWKPDYEPVPHVGHKTKRNTNRPODVKFGGQIVGRRGPPIPKARRPEGRWTW 998  
Db 2295 PVLITWKRCYEPPTVLG-----CALPPTPQ-----TPVPPRRRRRAKVL 2335  
Qy 999 AQPCYPM-----PLYGNKDR-RSTGKSWG 1021  
Db 2336 TDONVEGVLEMAKDVLSPLQDNNDSGHSTGADTG 2370

RESULT 12  
S68016  
ATPase/RNA helicase - hepatitis C virus (fragment)  
C;Species: hepatitis C virus  
C;Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: S68016  
R;Jin, L.; Peterson, D.L.  
Arch. Biochem. Biophys. 323, 47-53, 1995  
A;Title: Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA  
A;Reference number: S68016; MUID:96019946; PMID:7487072  
A;Accession: S68016  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-386 <JIN>  
A;Cross-references: UNIPROT:Q04045  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; nonstructural protein; nucleotide binding; P-loop; polyprotein  
F;24-31/Region: nucleotide-binding motif A (P-loop)  
F;86-91/Region: nucleotide-binding motif B  
F;90-93/Region: DEXH motif

Query Match 33.5%; Score 1980.5; DB 2; Length 386;  
Best Local Similarity 92.6%; Pred. No. 2.4e-116;  
Matches 376; Conservative 4; Mismatches 5; Indels 21; Gaps 2;  
Qy 246 PVFTDN-SSPPVVDQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLFGA 304  
Db 1 PVFTDNSSPPVVDQSFQVAHLHAPTSGSKTKVPAAYAAQGYKVLNPSVAATLFGA 60  
Qy 305 YMSKAHGIDPNIRTVITTTGSPITTYSTYKFLADGCGSGAYDIIICDCHSTDATSI 364  
Db 61 YMSKAHGV-----YKFLADGCGSGAYDIIICDCHSTDATSI 100  
Qy 365 LGIGTVLDQAETAGARLVVLTATPPGVSVPVPHNIEVALSTTGEIPFYKALPLEVIK 424

Db 101 LGIGTVLDQAETAGARLVVLTATPPGVSVPVPHNIEVALSTTGEIPFYKALPLEAIK 160  
Qy 425 GGRHLIFCHSKKKCDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTG 484  
Db 161 GGRHLIFCHSKKKCDELAALVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGFSG 220  
Qy 485 DFDVSIDCNTCVTQTVDFSLDPTFTTITLPODAVSRTOGRTGRGKGIYRFVAPGE 544  
Db 221 DFDVSIDCNTCVTQTVDFSLDPTFTTITLPODAVSRTOGRTGRGKGIYRFVAPGE 280  
Qy 545 RPSGMPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWEGVFTGLT 604  
Db 281 RPSGMPDSSVLCYDAGCAWYELTPAETTVRLRAYMNTPGLPVCQDHLFEWEGVFTGLT 340  
Qy 605 HIDAHLFSLQTKSGENLPLYLVAYQATVCARAQAPPPSWDQMKCL 650  
Db 341 HIDAHLFSLQTKSGENLPLYLVAYQATVCARAQAPPPSWDQMKCL 386

RESULT 13  
PC2219  
polypeptide - hepatitis C virus (type 5a) (fragments)  
N;Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A prote  
C;Species: hepatitis C virus  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C;Accession: PC2219  
R;Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.  
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994  
A;Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of the h  
A;Reference number: PC2219; MUID:94338342; PMID:7520237  
A;Accession: PC2219  
A;Molecule type: mRNA  
A;Residues: 1-876 <STU>  
A;Cross-references: UNIPROT:Q81242; GB:L29577; GB:L29578; GB:L29579  
A;Experimental source: serum  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: glycoprotein  
F;1-191/Product: core #status predicted <COB>  
F;68-78/Region: variable  
F;192-247/Product: E1 (carboxyl end) #status predicted <ERE>  
F;248-411/Product: E2/NS1 (amino end) #status predicted <ENR>  
F;248-338/Region: E2  
F;339-411/Region: NS1 (amino end)  
F;412-783/Product: NS3 #status predicted <NSR>  
F;784-837/Product: NS4A #status predicted <NSA>  
F;838-876/Product: NS4B #status predicted <NSB>  
F;281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 33.3%; Score 1966.5; DB 2; Length 876;  
Best Local Similarity 52.5%; Pred. No. 6.2e-115;  
Matches 426; Conservative 66; Mismatches 148; Indels 171; Gaps 18;  
Qy 69 SRKHGGPKDERHVGDLGNTADKQ-CVADVSIEDSVISLGDHCIIIGRTLHVHEKADDL 127  
Db 103 SRPMWGNDRPRKSRNKGKVIDTLTCGFAD-----LMGYIPLVGGPI----- 144  
Qy 128 GKGGNEBRTKTGNAGSLACGVIGIAQLN-----SGCNCISY-----PGH--- 168  
Db 145 -----GGVARALAHGVRLVDGVNATGNLPGCCSFSIFILALLSCLTVPASAPT 193  
Qy 169 -----ITGHRMAWK-----GSAARTTSQFVSLFAPGAKQNETHV 203  
Db 194 TALLVAQLLRIPQVVIDIAGSH--WGVLFAPAAAYASVANNWTKVVLVFLFAGVDATTQI 251  
Qy 204 TGGAAATTSGLTSLFSPGASQNIQLI-----VDPIPV-----ENLET----- 241  
Db 252 SGGSSAQTTTGIASTFIRGAQQLQLINTGSMHINRTALNCNDSLOTGFTAGFYHYKF 311  
Qy 242 -----TMRSPVFTD-----NSPVPVVDQSFQVAHLHAPTSGSKTKVPAAY 282  
Db 312 NSSCCPDMASCRALAFDQGWGTIYANISGPDSDPKYCHYPPRCG-----VVPAGE 366



```
QY 283 AAQGYKVLNPSVAATLFGAYMSKAHGIDPNIRTVRT-----ITTGSPIT 330
Db 367 VCGPVYCFPTSPVVGTGTD-----SKGH---PTYNMGSNVTDFLMMNTRPPIITTCASIT 418
QY 331 YSTYKFLADGGCGSGAYDIIICDECHSDATSIILGIGTVLDQAEATAGARLVVLATATPP 390
Db 419 YSTYKFLADGGCGSGAYDIIICDECHSQDATIILGIGTVLDQAEATAGARLVVLATATPP 478
QY 391 GSVTPHPNIEVALSTTGEIPYGYKAIPLFVTKGRHLIFCHSKKKKDELAKLVALGI 450
Db 479 GSVTPHPNIEVALPQEGEVFPYGRAIPLAFIKGRHLIFCHSKKKKDELAKQLTSLGV 538
QY 451 NAVAYRGLDVSIVPTSGDVVVVATDALMTGYGDFDSVIDCNTCTVQTVDLSLDPTEFTI 510
Db 539 NAVAYRGLDVAIVPTAGDVVVVCTDALMTGFTGDFDSVIDCNSAVTQTVDFSLDPTFTI 598
QY 511 ETITLQDVAVSRTQRRGTRGKPGIYRFVAPGERPSGMFDSVLCCEYDAGCAWYELTP 570
Db 599 ETITVQDVAVSRSQRRGTRGKPGIYRYVSAGERPSDMFDSVVLCEYDAGCAWYDLTP 658
QY 571 AETTIVRLRAYMNTPLPVCDHLEFWEQVFTGLTHIDAHFLSQTQSGENLPYLVAIQAT 630
Db 659 AETTIVRLRAYINTPLPVCDHLEFWEQVFTGLTHIDAHMLSQTQSGENFPYLVAIQAT 718
QY 631 VCARAQAPPSDQMWKCLIRLKPILHGPTELLYRLGAVQNEITLTHPVTKYIMTCSAD 690
Db 719 VCVRAKAPPSDQMTWKMLRLKPTLTGPTPLLYRLGVPQNEITLTHPIKTYIMACSAD 778
QY 691 LEVWISA-----CSGKPAIIPDREVLYREFDEMECS 722
Db 779 LEVITSTWLVGVVAAALAAAYCLTVGSVAIVGRILISGKPAIIPDREALYQQDFEMECS 838
QY 723 QHLPTYEQGMMLABQFKQKALG-LSRGKPA 752
Db 839 ASLPYMDETRAIAQGEKFKVLGPISTTGQA 869

RESULT 14
PS0326
polyprotein - hepatitis C virus (isolate Fla) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: PS0326
R:Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.
Gene 105, 167-172, 1991
A:Title: Two French genotypes of hepatitis C virus: homology of the predominant genotype
A:Reference number: PS0326; MUID:92039028; PMID:1718820
A:Accession: PS0326
A:Molecule type: genomic RNA
A:Residues: 1-492 <LiJ>
A:Cross-references: UNIPROT:Q9IFES; UNIPROT:O36579; UNIPROT:O36610; UNIPROT:Q03463; UNIPROT:M60220
A:Note: this sequence corresponds to nonstructural protein NS3 region
A:Note: translation of the nucleotide sequence is not complete
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

Query Match 25.1%; Score 1483.5; DB 2; Length 492;
Best Local Similarity 62.4%; Pred. No. 4.3e-85;
Matches 306; Conservative 22; Mismatches 63; Indels 99; Gaps 8;

QY 488 SVIDCNTCTVQTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPGERPS 547
Db 1 SVIDCNTCTVQTVDFSLDPTFTIETITLPQDAVSRTQRRGTRGKPGIYRFVAPGERPS 60
QY 548 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFWEQVFTGLTHID 607
Db 61 GMFDSSVLCCEYDAGCAWYELTPAETTVRLRAYMNTPLPVCDHLEFWEQVFTGLTHID 120
QY 608 AHFLSQTQSGENLPYLVAIQATVCARAQAPPSDQMWKCLIRLKPILHGPTELLYRLG 667
Db 121 AHFLSQTQSGENLPYLVAIQATVCARAQAPPSDQMWKCLIRLKPILHGPTELLYRLG 180
```

```
QY 668 AVQNEITLTHPVTKYIMTCSADLEVVTS-----ACS 699
Db 181 AVQNEITLTHPVTKYIMTCSADLEVVTSWLVGVVLAALAAAYCLSTGCVVIVGRVLS 240
QY 700 GKPAIIPDREVLYREFDEMECSQHLPIYEQGMMLABQFKQKALG-----SRGKPAIYP 755
Db 241 GKPAIIPDREVLYQSFDEMECSQHLPIYEQGMMLABQFKQKALGLLOLTASQAE-AITP 299
QY 756 DREVLYQQYD-----EMEECSQAAPYIBQAQVIAHQFKPKVGLGLINDQVVVTP---DKE 807
Db 300 AVQTNQWRLETFWAKHMNFIISGIQYLAGLSTLPG--NPATIASLMAFTAAVTSPLTTSOT 357
QY 808 ILYE-----AFDEMECASKAALIEEGORMAEMLKSKIQGLLG-----B45
Db 358 LLFNILGHWAAQLAAPGNAATAFVAGLAGAIGAIVGLGKVLVDILAGVAGVAGALVAF 417
QY 846 -----ILRRHVGPGEGAVQNMNRLIAPASRG 871
Db 418 KIMSGEVPSTEDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAVQNMNRLIAPASRG 477
QY 872 NHVSPTHYVP 881
Db 478 NHVSPTHYVP 487

RESULT 15
JQ1366
polyprotein - hepatitis C virus (French isolate) (fragments)
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: JQ1366
R:Krensdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.
J. Gen. Virol. 72, 2557-2561, 1991
A:Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implications
A:Reference number: JQ1366; MUID:92013977; PMID:1655961
A:Accession: JQ1366
A:Molecule type: genomic RNA
A:Residues: 1-716 <KRE>
A:Cross-references: UNIPROT:Q9PX22
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein; polyprotein
F:84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #state

Query Match 23.5%; Score 1386.5; DB 2; Length 716;
Best Local Similarity 46.6%; Pred. No. 8.8e-79;
Matches 336; Conservative 51; Mismatches 115; Indels 219; Gaps 27;

QY 200 ETHVTGGAARTTSGLTSLFSPCASQNIQLI-----VDFIPV---ENLETTMRSPVFT 249
Db 51 EYVTGSGTARTTQGLVSLFSRGAKQDIQLINTNGSWHINRTALNCNESLDTGWAGLFL- 109
QY 250 DNSSPVPVPSQFVAHLHAPTSGSKSTKV-----PAAYAAQGYKVLNPSVAATLFGA- 304
Db 110 -----YHKFNSSGCPERMASCRPLADFDQGM-----GPISYANGTPE 148
QY 305 -----YMSKAHGIDPNIRTVRTIT-----TGSPITVSTYVKFLA---DGCSCGGAY 348
Db 149 HRPYCHWHPKPCGIVP-----AQTVCGPVYCFPTSPVVVGTNKLGAPTYNMGCD--T 201
QY 349 DIII-----CDECHSDATSIILGIGTVLDQAEATAGARLVVLATATPP----- 390
Db 202 DVPVLNTRPPLGNWFGCTWNSSGFTKVCG-----APPCVIG 240
QY 391 -GSVTV-----PHNIEEVALSTTGIFP-----YGAIAPLEVIK-- 424
Db 241 AGNNTLYCPTDCFRKHP---EATYSCSGSPMTTPRCLVGPYVRLWHYPCTVNYTLFKVR 297
QY 425 ---GG---RHLIFCH--SKKCC---DELAACLVALGINAVAYRGLDVSIVPTSGDVVV 473
Db 298 MYGVGVEHRLQVACNWTGRERCNLDNRSELSPLLSITQW-----QVLPCS-----FT 347
QY 474 ATDALMTGYTGDPSVID-----C-NTCV 496
```



Db	348	TLPALTTGLIHLHQNIVDVQYLYGVSSIVSWAIKWYVILLFLLADARVCCLWNTCV	407
Qy	497	TQTVDFSLDPTFTIETITLQDQAVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSVLC	556
Db	408	TQTVDFSLDPTFTIETITLQDQAVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSVLC	467
Qy	557	ECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSQTQ	616
Db	468	ECYDAGCAWYELTPAETTVRLRAYMNTPGLPVCODHLEFWEGVFTGLTHIDAHFLSQTQ	527
Qy	617	SGENLPYLVAQATVCARAQAPPSWDMWKCLIRLAKPTLHGPTPLLYRLGAVQNEITLT	676
Db	528	SGENLPYLVAQATVCARAQAPPSWDMWKCLIRLAKPTLHGPTPLLYRLGAVQNEVILT	587
Qy	677	HPVTKYIMTMSADLEVVVISA-----CSGKPAIIPDR	708
Db	588	HPITKYIMTMSADLEVVVTSWTLVGGVLAALAAAYCLSTGCVIVGRVILSGKPAIIPDR	647
Qy	709	EVLRYREFDEMECSOHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLYQOY	764
Db	648	EVLRYREFDEMECSOHLPIYIEQGMMLAEQFKQKALGLLQTRSRQAE-VITPAVQTNQRL	706
Qy	765	D 765	
Db	707	E 707	

Search completed: November 7, 2005, 20:11:43  
Job time : 48.7031 secs

This page is blank.

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 20:01:16 ; Search time 140.844 Seconds  
(without alignments)  
3995.746 Million cell updates/sec

Title: US-10-658-782-6  
Perfect score: 5912  
Sequence: 1 MATKAVCVLKGDPVQGIIN.....GNKDRRSTGKSWKPGYPWP 1099

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2909.5	49.2	2436	2	Q81756	Q81756 hepatitis c
2	2909.5	49.2	3011	1	POLG_HCV1	P26664 h genome po
3	2905.5	49.1	3011	2	Q31FES	Q31FES hepatitis c
4	2875	48.6	3011	2	Q36579	Q36579 hepatitis c
5	2868	48.5	3011	2	Q36608	Q36608 hepatitis c
6	2868	48.5	3011	2	Q36610	Q36610 hepatitis c
7	2868	48.5	3015	2	Q9PMW9	Q9PMW9 hepatitis c
8	2868	48.5	3015	2	Q9PMW5	Q9PMW5 hepatitis c
9	2866	48.5	2908	2	Q61X04	Q61X04 hepatitis c
10	2861	48.4	3011	2	Q36609	Q36609 hepatitis c
11	2851	48.2	3011	2	Q36LS8	Q36LS8 hepatitis c
12	2850	48.2	3011	2	Q34633	Q34633 hepatitis c
13	2835	48.0	3011	2	Q3DIT6	Q3DIT6 hepatitis c
14	2821	47.7	3011	1	POLG_HCVH	P27958 h genome po
15	2783.5	47.1	3010	2	Q9J3H0	Q9J3H0 hepatitis c
16	2777.5	47.0	3010	2	Q9DTE8	Q9DTE8 hepatitis c
17	2775	46.9	3010	2	Q9J3G4	Q9J3G4 hepatitis c
18	2770	46.9	3013	2	Q9QNC0	Q9QNC0 hepatitis c
19	2768.5	46.8	3010	1	POLG_HCVJT	Q00269 h genome po
20	2766	46.8	3014	2	Q9DTE0	Q9DTE0 hepatitis c
21	2764.5	46.8	3010	2	Q9J3G8	Q9J3G8 hepatitis c
22	2764	46.8	3010	2	Q9J3G2	Q9J3G2 hepatitis c
23	2763.5	46.7	3010	2	Q9QIY2	Q9QIY2 hepatitis c
24	2763	46.7	3010	2	Q9QIY5	Q9QIY5 hepatitis c
25	2763	46.7	3010	2	Q9QIY6	Q9QIY6 hepatitis c
26	2762.5	46.7	3010	2	Q9QIX2	Q9QIX2 hepatitis c
27	2762.5	46.7	3010	2	Q9QIY1	Q9QIY1 hepatitis c
28	2761	46.7	3010	2	Q9DTE6	Q9DTE6 hepatitis c
29	2759	46.7	3010	2	Q9DTE9	Q9DTE9 hepatitis c
30	2759	46.7	3010	2	Q9J3H8	Q9J3H8 hepatitis c
31	2758	46.7	3010	2	Q81757	Q81757 hepatitis c

RESULT 1

ID	Q81756	PRELIMINARY;	PRT;	2436 AA.
AC	Q81756;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Polyprotein (Fragment)			
OS	Hepatitis C virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11103;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Choo Q.-L., Richman K., Han J.;			
RL	Submitted (MAY-1990) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M32084; AAA45677.1; -			
DR	PIR; PS0326; PS0326.			
DR	PIR; PS0327; PS0327.			
DR	PIR; PS0328; PS0328.			
DR	HSSP; P27958; 1A1V.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0019028; C:viral capsid; IEA.			
DR	GO; GO:0019031; C:viral envelope; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.			
DR	GO; GO:0003723; F:RNA binding; IEA.			
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.			
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.			
DR	GO; GO:0005198; F:structural molecule activity; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	GO; GO:0019079; P:transcription; IEA.			
DR	GO; GO:0019087; P:viral genome replication; IEA.			
DR	InterPro; IPR000345; CytC_heme_BS.			
DR	InterPro; IPR001410; DEAD.			
DR	InterPro; IPR011545; DEAD/DEAH_N.			
DR	InterPro; IPR002531; HCV NS1.			
DR	InterPro; IPR000745; HCV NS4a.			
DR	InterPro; IPR001490; HCV NS4b.			
DR	InterPro; IPR002868; HCV NS5a.			
DR	InterPro; IPR002166; HCV RdRP.			
DR	InterPro; IPR001650; Helicase_C.			
DR	InterPro; IPR004109; Peptidase_S29.			
DR	InterPro; IPR002518; Pept Ser Cys.			
DR	InterPro; IPR007095; RNA_pol_DS_PS.			
DR	InterPro; IPR007094; RNA_pol_Psvir.			
DR	Pfam; PF01560; HCV NS1; 1.			
DR	Pfam; PF01538; HCV NS2; 1.			
DR	Pfam; PF02907; HCV NS3; 1.			
DR	Pfam; PF01006; HCV NS4a; 1.			
DR	Pfam; PF01001; HCV NS4b; 1.			
DR	Pfam; PF01506; HCV NS5a; 1.			

32	2757.5	46.6	3010	2	Q9J3G5	Q9J3G5 hepatitis c
33	2757.5	46.6	3010	2	Q9QIX1	Q9QIX1 hepatitis c
34	2757.5	46.6	3010	2	Q9QIY3	Q9QIY3 hepatitis c
35	2756	46.6	3010	2	Q68788	Q68788 hepatitis c
36	2756	46.6	3010	2	Q9J3H5	Q9J3H5 hepatitis c
37	2756	46.6	3010	2	Q9QIX6	Q9QIX6 hepatitis c
38	2756	46.6	3013	2	Q6J6P5	Q6J6P5 hepatitis c
39	2755.5	46.6	3010	2	Q81541	Q81541 hepatitis c
40	2755	46.6	3010	2	Q9QIY9	Q9QIY9 hepatitis c
41	2755	46.6	3010	2	Q9QIY9	Q9QIY9 hepatitis c
42	2755	46.6	3010	2	Q9WMX2	Q9WMX2 hepatitis c
43	2754.5	46.6	3010	2	Q9J3I1	Q9J3I1 hepatitis c
44	2754	46.6	1984	2	Q7T4V8	Q7T4V8 hepatitis c
45	2753.5	46.6	3010	2	Q807P3	Q807P3 hepatitis c

DR	Pfam; PF00271; Helicase C; 1.	
DR	Pfam; PF00998; Viral RdRP; 1.	
DR	SMART; SM00487; DEXD; 1.	
KW	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.	
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;	
KW	Polyprotein; Transmembrane.	
FT	NON_TER 1 1	
FT	NON_TER 2436 2436	
SQ	SEQUENCE 2436 AA; 264734 MW; D7B9872900BE3125 CRC64;	
Query Match 49.2%; Score 2909.5; DB 2; Length 2436;		
Best Local Similarity 53.3%; Pred. No. 7.9e-166;		
Matches 622; Conservative 33; Mismatches 92; Indels 419; Gaps 16;		
QY	224 SONIQLIVDFIPVENLETTMRSPVFTDNSSPPVQSFQVAHLHAPTGSGSKTKVPAAYA	283
DB	2736 TRGVAKAVDFIPVENLETTMRSPVFTDNSSPPVQSFQVAHLHAPTGSGSKTKVPAAYA	795
QY	284 AQGYKVLVLPNSVAATLGFCAYSKAHGDIPNIRTGVRTITTTGSPITYSTYGRFLADGGC	343
DB	796 AQGYKVLVLPNSVAATLGFCAYSKAHGDIPNIRTGVRTITTTGSPITYSTYGRFLADGGC	855
QY	344 SGAYDIIICDECHSDTATSIIGITVLDQAEAGARLVVLTATPPGCVTVPHNIEEV	403
DB	856 SGAYDIIICDECHSDTATSIIGITVLDQAEAGARLVVLTATPPGCVTVPHNIEEV	915
QY	404 ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKDELAALVALGINAVAYYRGLDVSU	463
DB	916 ALSTTGEIPYGAIPLEVIKGRHLIFCHSKKDELAALVALGINAVAYYRGLDVSU	975
QY	464 IPTSGDVVVVATDALMTGYTGFDSVIDNCNTCTVTQTVDFSLDPTFTIETITLPPQDAVSRT	523
DB	976 IPTSGDVVVVATDALMTGYTGFDSVIDNCNTCTVTQTVDFSLDPTFTIETITLPPQDAVSRT	1035
QY	524 QRRGRTGRGKPGIYRFVAPGERSGMPDSSVLCEDYDAGCAYVELTPAETTVRLAYMNT	583
DB	1036 QRRGRTGRGKPGIYRFVAPGERSGMPDSSVLCEDYDAGCAYVELTPAETTVRLAYMNT	1095
QY	584 PGLPVQCDHLEFEGVFTGLTHDAFLSOTKSGENLPYLVAQVATVCARAQAPPSWD	643
DB	1096 PGLPVQCDHLEFEGVFTGLTHDAFLSOTKSGENLPYLVAQVATVCARAQAPPSWD	1155
QY	644 QMWKILIRLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTQMSADLEVVTTS	696
DB	1156 QMWKILIRLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTQMSADLEVVTTS	1215
QY	697 -----ACSGKPAIIPDREVLRYREDFEDEMBCSQHLPIYEQGMMLA	735
DB	1216 VLAALAAAYCLSTGCVVIGRWLSGKPAIIPDREVLRYREDFEDEMBCSQHLPIYEQGMMLA	1275
QY	736 EQFKQKALG-----SRGKPAIIPDREVLRYREDFEDEMBCSQHLPIYEQGMMLA	786
DB	1276 EQFKQKALG-----SRGKPAIIPDREVLRYREDFEDEMBCSQHLPIYEQGMMLA	1333
QY	787 FKEVILGLIDNDQVWVTP-----DKEILYE-----AFDEMEECASKAALI	826
DB	1334 -NPAIASLMAFTAAVTSPLTSTQTLNFIILGWVAQAAPGAATFVAGLAGAIGSV	1392
QY	827 EEOQMAEMLSKIQGLG-----	847
DB	1393 GLGKFLIDILAGYAGVAGVAFKIMSGEVSPSTEDLVNLLPAILSPGALVWGVCAAIL	1452
QY	848 RRVHVGEGAVQVNNRLIAFASRGNHVSPTHYVPS	882
DB	1453 RRVHVGEGAVQVNNRLIAFASRGNHVSPTHYVPSDAAARVTAISSLVTQLLRLHQ	1512
QY	883 -----	882
DB	1513 WISSECTTPCSGSLRDINDWICEVLSDFKTLKAKLMPQLPGIPFVSCQGYKGVWRVD	1572
QY	883 -----	882
DB	1573 GIMHTRCHGAEITGHVKNGTMRIVGPRTCRNMMSGTFPINAYTGTGPTPLPAPNYTAL	1632

QY	883 -----RSTRFA-----	888
DB	1633 WRVSAEYVEIRQVGDHYVTGTTDNLKCPCQVPSPEFFTELDGVLRHFAFPCKPLLR	1692
QY	889 -----QALPVMARPDY-----NPP-----	902
DB	1693 EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSLMTDPSHTAEAGRRRLARGSPSVASS	1752
QY	903 -----LVET---WKK-----PDYEPVVH	918
DB	1753 ASOLSAPSLKATCTANHSDPAELIEANLLWRQEMGNTRVSEKNKVLDSFDPLVAE	1812
QY	919 G-----RSSRRFAQALPVMARPDYNNPPLVETWKKPDYEPVVHGRKTKRNTNR	966
DB	1813 EDEREISVPAELTKRRRFAQALPVMARPDYNNPPLVETWKKPDYEPVVH-----	1862
QY	967 RPDVKFPGGQIVGRGPPPKARR	992
DB	1863 -----GCPLPPPKSPVPPPRK	1879
RESULT 2		
POLG_HCV1	STANDARD; PRT; 3011 AA.	
ID	POLG_HCV1	
AC	P26664;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	25-OCT-2004 (Rel. 45, Last annotation update)	
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);	
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2	
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)	
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)	
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein	
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein	
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48) [.	
OS	Hepatitis C virus (isolate 1) (HCV).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus.	
OX	NCBI_TaxID=11104;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=91172826; PubMed=1848704;	
RA	Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,	
RA	Gallagos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,	
RA	Bradley D.W., Kuo G., Houghton M.	
RT	"Genetic organization and diversity of the hepatitis C virus.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455 (1991).	
CC	-I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are	
CC	hydrophobic, suggesting a possible membrane-related function. NS3	
CC	and NS5 may play a role in the viral RNA replication.	
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral	
CC	precursor polyprotein, commonly with Asp or Glu in the P6	
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.	
CC	-I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +	
CC	{RNA} (N).	
CC	-I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a	
CC	lipoprotein envelope. The envelope consists of two proteins:	
CC	protein M and glycoprotein E. The nucleocapsid is a complex of	
CC	protein C and mRNA.	
CC	-I- SIMILARITY: Contains 1 peptidase S29 domain.	
CC	-I- SIMILARITY: Contains 1 peptidase U39 domain.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ).	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .	
CC	-----	
DR	EMBL; M62321; AAA45676.1; -.	
DR	FIR; A39166; GNWVC3.	

DR PDB; 1HEI; X-ray; A/B=1206-1656.  
 DR PDB; 1ONB; NMR; A=1349-1507.  
 DR MEROPS; S29.001; -  
 DR MEROPS; U39.001; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.  
 DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 DR InterPro; IPR004109; Peptidase\_S23.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01535; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR 3D-structure; ATP-binding; Coat protein; Core protein;  
 KW Envelope protein; Glycoprotein; Helicase; Hydrolase;  
 KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 KW Serine protease; Transferase; Transmembrane.  
 FT INIT\_MBT 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3011  
 FT TRANSMEM 347 369  
 FT ACT\_SITE 1083 1083  
 FT ACT\_SITE 1107 1107  
 FT ACT\_SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 196 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 476 476  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2364 2364  
 FT CARBOHYD 2789 2789  
 FT STRAND 1224 1226  
 PDB; 1HEI; X-ray; A/B=1206-1656.  
 PDB; 1ONB; NMR; A=1349-1507.  
 MEROPS; S29.001; -  
 MEROPS; U39.001; -  
 InterPro; IPR001410; DEAD.  
 InterPro; IPR002522; HCV\_capsid.  
 InterPro; IPR002521; HCV\_core.  
 InterPro; IPR002519; HCV\_env.  
 InterPro; IPR002531; HCV\_NS1.  
 InterPro; IPR000745; HCV\_NS4a.  
 InterPro; IPR001490; HCV\_NS4b.  
 InterPro; IPR002868; HCV\_NS5a.  
 InterPro; IPR002166; HCV\_RdRP.  
 InterPro; IPR001650; Helicase\_C.  
 InterPro; IPR009003; Pept\_Ser\_Cys.  
 InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 InterPro; IPR004109; Peptidase\_S23.  
 InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 InterPro; IPR007094; RNA\_pol\_PSVir.  
 Pfam; PF01543; HCV\_capsid; 1.  
 Pfam; PF01542; HCV\_core; 1.  
 Pfam; PF01535; HCV\_env; 1.  
 Pfam; PF01560; HCV\_NS1; 1.  
 Pfam; PF01538; HCV\_NS2; 1.  
 Pfam; PF02907; HCV\_NS3; 1.  
 Pfam; PF01006; HCV\_NS4a; 1.  
 Pfam; PF01001; HCV\_NS4b; 1.  
 Pfam; PF01506; HCV\_NS5a; 1.  
 Pfam; PF00271; Helicase\_C; 1.  
 Pfam; PF00998; Viral\_RdRP; 1.  
 SMART; SM00487; DEXDC; 1.  
 3D-structure; ATP-binding; Coat protein; Core protein;  
 Envelope protein; Glycoprotein; Helicase; Hydrolase;  
 Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 Serine protease; Transferase; Transmembrane.  
 INIT\_MBT 1 1  
 CHAIN 1 115  
 CHAIN 116 191  
 CHAIN 192 383  
 CHAIN 384 729  
 CHAIN 730 1006  
 CHAIN 1007 1615  
 CHAIN 1616 1862  
 CHAIN 1863 2013  
 CHAIN 2014 3011  
 TRANSMEM 347 369  
 ACT\_SITE 1083 1083  
 ACT\_SITE 1107 1107  
 ACT\_SITE 1165 1165  
 NP\_BIND 1230 1237  
 SITE 1316 1319  
 CARBOHYD 196 196  
 CARBOHYD 209 209  
 CARBOHYD 234 234  
 CARBOHYD 305 305  
 CARBOHYD 417 417  
 CARBOHYD 423 423  
 CARBOHYD 430 430  
 CARBOHYD 448 448  
 CARBOHYD 476 476  
 CARBOHYD 532 532  
 CARBOHYD 540 540  
 CARBOHYD 556 556  
 CARBOHYD 576 576  
 CARBOHYD 623 623  
 CARBOHYD 645 645  
 CARBOHYD 2041 2041  
 CARBOHYD 2077 2077  
 CARBOHYD 2240 2240  
 CARBOHYD 2364 2364  
 CARBOHYD 2789 2789  
 STRAND 1224 1226

Query Match 49.2%; Score 2909.5; DB 1; Length 3011;

Best Local Similarity 53.3%; Pred. No. 1.1e-165; Mismatches 92; Indels 419; Gaps 16;

Matches 622; Conservative 33; Mismatches 92; Indels 419; Gaps 16;

QY 224 SQNIQLIVDFIPVENLETMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAVA 283

Db 1186 TRGVAKAVDFIPVENLETMRSPVFTDNSSPPVPPQSFQVAHLHAPTSGSKSTKVPAAVA 1245

QY 284 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVGRTITTTGSPITTYGKFLADGGC 343

Db 1246 AQGYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTVGRTITTTGSPITTYGKFLADGGC 1305

QY 344 SGGAYDIIICDECHSTDATSIIGICTVLDQAEATAGARLVVLATATPPGSVTVPHPNIBV 403

Db 1306 SGGAYDIIICDECHSTDATSIIGICTVLDQAEATAGARLVVLATATPPGSVTVPHPNIBV 1365

QY 404 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELAALKVALGINAVAYYRGLDVSV 463

Db 1366 ALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELAALKVALGINAVAYYRGLDVSV 1425

SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;

```
Qy 464 IPTSGDVVVVATDALMTGYTGDVDSVIDNCNTCVTQVDFSLDPTFTTETITLPQDAVSRT 523
Db 1426 IPTSGDVVVVATDALMTGYTGDVDSVIDNCNTCVTQVDFSLDPTFTTETITLPQDAVSRT 1485
Qy 524 ORRGTRGKPGIYRFVAPGERSGMDSSVLCYCYDAGCAWYELTETVRLRAYMNT 583
Db 1486 ORRGTRGKPGIYRFVAPGERSGMDSSVLCYCYDAGCAWYELTETVRLRAYMNT 1545
Qy 584 PGLPVCODHLEFWEFVGTGLTHIDAHFLSOTKQSGENLPYLVAQVATVCARAQAPPPSWD 643
Db 1546 PGLPVCODHLEFWEFVGTGLTHIDAHFLSOTKQSGENLPYLVAQVATVCARAQAPPPSWD 1605
Qy 644 QMVKCLIRLKLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS 696
Db 1606 QMVKCLIRLKLPTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSVLVGG 1665
Qy 697 -----ACSGKPAIIPDREVLVREDEMEECQHLPIYIEQGMMLA 735
Db 1666 VLAALAAAYCLSTGCVTVGRVVLGSKPAIIPDREVLVREDEMEECQHLPIYIEQGMMLA 1725
Qy 736 EQFKQKALGL-----SRGKPAIVDPKEVLYQOYD-----EWEECQAAPYIEQAQVIAHQ 786
Db 1726 EQFKQKALGLLOTASRAE-VIAPAVQTNQKLETFWAKHWNFISGIQVLAGLSTLPG- 1783
Qy 787 FKSKVLGLINDQVVVTP-----DKEILYE-----AFDEMEECASKAALI 826
Db 1784 -NPAIASMAFTAATAVTSPLTTSQTLLFNILGWVAAQAAPGAATAFVGAGLAGAAGISV 1842
Qy 827 EEQRMMAEMLKSKIQGLG-----IL 847
Db 1843 GLGKVLIDILAGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPALSPGALWGVGCAAIL 1902
Qy 848 RRHVGPGEQAVQWNNRLIAFASRGNHVSPTHYVPS-----882
Db 1903 RRHVGPGEQAVQWNNRLIAFASRGNHVSPTHYVPSDAAARVAILSSLTQVLLRLHQ 1962
Qy 883 -----882
Db 1963 WISSECTTPCSGSLRDINDWICEVLSDFKTLKAKLMPQLPGIPFVSCQGYKGVWRVD 2022
Qy 883 -----882
Db 2023 GIMHTRCHGAEITHVKNGTMRIVGPRTCRNMWSTGFPINAYVTGCTELPAPNVTAL 2082
Qy 883 -----RSRFEA-----888
Db 2083 WRVSAEYVEIRQVGDHYVTGTTDNLCPCQVPSEFFTELDGVRLHRFAPCPKPLLR 2142
Qy 889 -----QALPWVARPDY-----NPP-----902
Db 2143 EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPHSITAEAAAGRLARGSPFSVASS 2202
Qy 903 -----LVEF---WKK-----PYEPPVTH 918
Db 2203 ASQLSAPSLKATCANHDSPEALIEANLLWRQEMGNIITRVSEKNKWLIDSFDELVAE 2262
Qy 919 G-----RSSRRFAQALPVWARPDPNPLVETWKKPDYEPVPPVHGRKTKRNTNR 966
Db 2263 EDREISVPAEILKSRFAQALPVWARPDPNPLVETWKKPDYEPVPPVH-----2312
Qy 967 RPODVKPPGGQIVGRGGPIPKARR 992
Db 2313 -----GCPLPPKSPVPPPPRK 2329
```

## RESULT 3

```
Q9IF5 PRELIMINARY; PRT; 3011 AA.
AC Q9IF5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polyprotein.

Query Match 49.1%; Score 2905.5; DB 2; Length 3011;
Best Local Similarity 53.3%; Pred. No. 1.9e-165;
Matches 621; Conservative 34; Mismatches 92; Indels 419; Gaps 16;

Qy 224 SQNIQLIVDFIPVENLETMRSPVTDNSSPPVQSFQVAHLHAPTSGSKTKVPAAYA 283
Db 1186 TRGKAVKAVDFIPVENLETMRSPVTDNSSPPVQSFQVAHLHAPTSGSKTKVPAAYA 1245
```

```
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21262212; PubMed=11369872;
RA Lanford R.E., Lee H., Chavez D., Guerra B., Brasky K.M.;
RT "Infectious cDNA clone of the hepatitis C virus genotype 1 prototype
sequence.";
RL J. Gen. Virol. 82:1291-1297(2001).
DR EMBL; AF271632; AAF81759.1; -.
DR PIR; A44150; A44150.
DR PIR; Q0804; Q0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSP; QBUYS1; ICWX.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_env.
DR InterPro; IPR002519; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept Ser Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV NS4a; 1.
DR Pfam; PF01001; HCV NS4b; 1.
DR Pfam; PF01506; HCV NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327126 MW; 2489CE74AC864E58 CRC64;
```

284 AQGYKVLVNPSTVAATLGFAGYMSKAHGIDPNIRTTGRTTITGSPITTYSTYTGFLADGGC 343  
1246 AQGYKVLVNPSTVAATLGFAGYMSKAHGIDPNIRTTGRTTITGSPITTYSTYTGFLADGGC 1305  
344 SGGAYDIIICDECHSTDATSILGIGTGTVDQAEAGARLVLAATATPGSVTVPHNLEEV 403  
1306 SGGAYDIIICDECHSTDATSILGIGTGTVDQAEAGARLVLAATATPGSVTVPHNLEEV 1365  
404 ALSTTGIPYGYKAIPLEVIKGRHLLIFCHSKKKDELAALVALGINAVAYRGLDVSV 463  
1366 ALSTTGIPYGYKAIPLEVIKGRHLLIFCHSKKKDELAALVALGINAVAYRGLDVSV 1425  
464 IPTSGDVVVVATDALMTGYTGFDSVIDCNTVQTVDVDFSLDPTFTTITLPODAVSRT 523  
1426 IPTSGDVVVVATDALMTGYTGFDSVIDCNTVQTVDVDFSLDPTFTTITLPODAVSRT 1485  
524 ORRGTRGRGPGIYRFVAPGERPSGMDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT 583  
1486 ORRGTRGRGPGIYRFVAPGERPSGMDSSVLCECYDAGCAWYELTPAETTVRLRAYMNT 1545  
584 PGLPVCODHLEFMEGVFTGLTHIDAHFLSOTKQSGENLPYLVAQATVCARAQAPPSWD 643  
1546 PGLPVCODHLEFMEGVFTGLTHIDAHFLSOTKQSGENLPYLVAQATVCARAQAPPSWD 1605  
644 QMWKCLIRLKPRTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTS----- 696  
1606 QMWKCLIRLKPRTLHGPTPLLYRLGAVQNEITLTHPVTKYIMTCSADLEVVTSWLVVGG 1665  
697 -----ACSGKPAIPDRVLYREDEMECSQHLPIYEQGMWLA 735  
1666 VLAALAAAYCLSTGCWIVGRVLSGKPAIPDRVLYREDEMECSQHLPIYEQGMWLA 1725  
736 EQFKQKALGL-----SRGKPAIPDRVLYREDEMECSQHLPIYEQGMWLA 786  
1726 EQFKQKALGLLQTSARQAE-VIAFVQTNQWKLTFWAKHWNFISGTYLAGLSTLPG- 1783  
787 FKEKVLGLINDQVVVTP-----DKBILYE-----AFDEMECSKAAALI 826  
1784 -NPAISLMAFTAATVSPLTSTQTLFNILGWVAAQAAPGAATAFVAGLAGAAGTGV 1842  
827 EBGORMAEMLSKTKQGLLG-----IL 847  
1843 GLGKVLIDILAGYGAGVAGALVAFKIMSGEVSPSTEDLVNLLPAILSPGALVGVVCAAIL 1902  
848 RRHVGPGEVQVNMNRLIAFASRGNHVSPTHYVPS----- 882  
1903 RRHVGPGEVQVNMNRLIAFASRGNHVSPTHYVPSDDAAARVAILSSLTQTLRLRHQ 1962  
883 ----- 882  
1963 WISSECTPCSGMLRDIWDWICEVLSDFKTWLKAQKMLPQLPGIPFVSCQRYKGVWRGD 2022  
883 ----- 882  
2023 GIMHTRCHGAEITGHVKNQGMTRIVGPRTCRNMWSGTFFINAYTGPCTPLPAPNYTFAL 2082  
883 -----RSRREA----- 888  
2083 WRVSAEEVEIRQVGDHFHYVTGMTDNLKPCQVPSPEFFTELDGRLHRFAPCKPLLR 2142  
889 -----QALPVWARPDY-----NPP----- 902  
2143 EEVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLTDPSSHITAEAGRRRLARGSPPSVASS 2202  
903 -----LVET-----WKK-----PDVEPPVH 918  
2203 ASQLSAPSLKATCTANHDSPDALTEANLLWRQEMGNITRVESKNKVVILDSFDPLVAE 2262  
919 G-----RSSRRFAQALPVWARPDPNPLVETWKKPDYPPPPVHGRKTKRNTNR 966  
2263 EDEREISVPAELIKRSRFAQALPVWARPDPNPLVETWKKPDYPPPPVH----- 2312  
967 RPQDVKFPGGQIVGRRGPPIPKARR 992

Db 2313 -----GCPLPPPKSPVPPPRK 2329  
RESULT 4  
O36579 PRELIMINARY; PRT; 3011 AA.  
AC O36579; SEQUENCE FROM N.A.  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Polypeptide.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97373636; PubMed=9228008; DOI=10.1126/science.277.5325.570;  
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,  
RA Rice C.M.;  
RT "Transmission of hepatitis C by intrahepatic inoculation with  
RT "Transcribed RNA.";  
RL Science 277:570-574(1997).  
DR EMBL; AF009606; AAB6324.1; -.  
DR PIR; A44150; A44150.  
DR PIR; PQ0804; PQ0804.  
DR PIR; PS0326; PS0326.  
DR PIR; PS0327; PS0327.  
DR PIR; PS0328; PS0328.  
DR PDB; INLL; X-ray; A/B=1017-1214.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003668; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:000350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH N.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4a.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF02907; HCV\_NS3; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.

KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; Transmembrane.	
KW	SEQUENCE 3011 AA; 327184 MW; E250EB809C63C1B9 CRC64;	
SQ	Query Match 48.6%; Score 2875; DB 2; Length 3011; Best local similarity 51.8%; Pred. No. 1.3e-163; Matches 617; Conservative 41; Mismatches 104; Indels 430; Gaps 17;	
QY	204 TGGAAATTTSGLTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDSSPPVQSFQ 262	
DB	1165 SGGLPLCPAGHAVGLFRAAVCTRGVAKAVDFIPVENLETTMRSPVFTDSSPPAVQSFQ 1224	
QY	263 VAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGIDPNIRTGVRT 322	
DB	1225 VAHLHAPTSGSKTKVPAAYAAQGYKVLVLPNSVAATLGFAYMSKAHGVDPNIRTGVRT 1284	
QY	323 ITTGSPIITYSTYKFLADGCGSGAYDIIICDECHSTDATSIIGITGVLDQAGTAGARLV 382	
DB	1285 ITTGSPIITYSTYKFLADGCGSGAYDIIICDECHSTDATSIIGITGVLDQAGTAGARLV 1344	
QY	383 VLATATPPGSVTPHPNIEBALSTTGEIPFYKAIPILEVIKGRHLIFCHSKKKDELA 442	
DB	1345 VLATATPPGSVTWSHPNIEBALSTTGEIPFYKAIPILEVIKGRHLIFCHSKKKDELA 1404	
QY	443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGDFSDVIDCNTCVQTQVDF 502	
DB	1405 AKLVALGINAVAYRGLDVSIVPTSGDVVVVSTDALMTGYTGDFSDVIDCNTCVQTQVDF 1464	
QY	503 SLDPTFTIETITLPQDAVSTQRRGTGRGKPGIYRFVAPGERPSGMFDSVLCVCYDAG 562	
DB	1465 SLDPTFTIETITLPQDAVSTQRRGTGRGKPGIYRFVAPGERPSGMFDSVLCVCYDAG 1524	
QY	563 CAMELTTPAETTVRLRAYMNTPLGVQODHLEFWEGVFTGLTHIDAHFLSOTKSGENLP 622	
DB	1525 CAMELTTPAETTVRLRAYMNTPLGVQODHLEFWEGVFTGLTHIDAHFLSOTKSGENFP 1584	
QY	623 YLVAYQATVCARAQAAPPPSDQMWKCLIRLKTPLHGTPTLLYRLGAVQNEITITHPVTKY 682	
DB	1585 YLVAYQATVCARAQAAPPPSDQMWKCLIRLKTPLHGTPTLLYRLGAVQNEITITHPVTKY 1644	
QY	683 IMTCMSADLEWVTS-----ACSGKPAIIPREVLYRE 714	
DB	1645 IMTCMSADLEWVTSYVWLVGGVLAALAAAYCLSTGCVVIVGRIVLSGKPAIIPREVLYQE 1704	
QY	715 FDEMECSQHLPIEQGMMLAEQFKOKALGL-----SRGKPAIYVDPKEVLYQYQVDEMEEC 770	
DB	1705 FDEMECSQHLPIEQGMMLAEQFKOKALGLLQTSRQAE-VITPAVQTNWKL-EVFWA 1762	
QY	771 SQAPYIEQAQVIAHQFKEKVLGLINDQVV-----VTPDKELIYE----- 811	
DB	1763 KHMWNFISGIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLTTGTLLFNILGGWV 1816	
QY	812 -----AFDEMECASKAALIEEGQRMALMSKIQGLLG----- 845	
DB	1817 AAQLAAPGAATAFVAGLAGAAGISVGLGVLDILAGYAGVAGALVAFKIMSGEVPST 1876	
QY	846 -----ILRRHHVPGEGAVQWMNRLIAFASRGNHVSPPHYVP 881	
DB	1877 EDLVNLLPALSPCALVGVVCAAILRRHHVPGEGAVQWMNRLIAFASRGNHVSPPHYVP 1936	
QY	882 S----- 882	
DB	1937 ESDAAARVTAISLTVTQLRLRLHQLWISSECTTPCSGSMWRDWDWICEVLSDFKTWLK 1996	
QY	883 ----- 882	
DB	1997 AKLMPQLPGIPFVSCQRYGVWRGDGINHTRCHGAEITGHVKNGTMRIVGPRTCRNMW 2056	
QY	883 ----- 882	
DB	2057 SGTFFPINAYTGTCTPLPAPNYKFAWRVSAEYVEIRRVGDFHYSGMTTDLNKCPCQI 2116	
QY	883 -----RSRRFA-----QALPVWARPDPY----- 899	

DB	2117 PSPEFFTELDGVRHLRHFADPKPCLLRREEVSRVGLHEYPVGSQLPCEPDPVAVLTSMLT 2176	
QY	900 -----NPP-----LVET---WKK- 909	
DB	2177 DPSHITAEAGRRRLARGSPSSMASSASOLSAPSLKATCTANHDSPEALIEANILLWROE 2236	
QY	910 -----PDYEPVVHG-----RSSRRFAQALFVWARPDPYNNP 940	
DB	2237 MGNITRVESENKVVILDSFDPLVAEEDEREVSVAEILRKSRFFARALPVWARPDPYNNP 2296	
QY	941 LVETWKKPDYEPVVHGRKTKNTNRPPQDVKFPGGQIVGRRGPPPIPKARR 992	
DB	2297 LVETWKKPDYEPVVH-----GCPLPPRSPVPVPPRK 2329	
RESULT 5		
O36608	PRELIMINARY; PRT; 3011 AA.	
ID	AC O36608; (TEMBLrel. 05, Created)	
AC	01-JAN-1998 (TEMBLrel. 05, Last sequence update)	
DT	01-MAR-2004 (TEMBLrel. 26, Last annotation update)	
DE	Polyprotein.	
OS	Hepatitis C virus strain H77.	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.	
OX	NCBI_TaxID=63746;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H77;	
RX	MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;	
RA	Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;	
RT	"Transcripts from a single full-length cDNA clone of hepatitis C virus	
RT	are infectious when directly transfected into the liver of a	
RT	chimpanzee."	
RL	Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743(1997).	
DR	EMBL; AF011751; AAB67036.1; --	
DR	PIR; A44150; A44150.	
DR	PIR; PQ0804; PQ0804.	
DR	PIR; PS0326; PS0326.	
DR	PIR; PS0327; PS0327.	
DR	PIR; PS0328; PS0328.	
DR	HSSP; P27958; 1HEI.	
DR	GO; GO:0019028; C:vital capsid; IEA.	
DR	GO; GO:0019031; C:vital envelope; IEA.	
DR	GO; GO:0005224; F:ATP binding; IEA.	
DR	GO; GO:0008026; F:ATP-dependent helicase activity; IEA.	
DR	GO; GO:0003723; F:RNA binding; IEA.	
DR	GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.	
DR	GO; GO:0008236; F:serine-type peptidase activity; IEA.	
DR	GO; GO:0005198; F:structural molecule activity; IEA.	
DR	GO; GO:0005508; P:proteolysis and peptidolysis; IEA.	
DR	GO; GO:0006350; P:transcription; IEA.	
DR	GO; GO:0019079; P:vital genome replication; IEA.	
DR	GO; GO:0019087; P:vital transformation; IEA.	
DR	InterPro; IPR000345; CytC heme_BS.	
DR	InterPro; IPR001410; DEAD.	
DR	InterPro; IPR011545; DEAD/DEAH_N.	
DR	InterPro; IPR002522; HCV_capsid.	
DR	InterPro; IPR002521; HCV_core.	
DR	InterPro; IPR002519; HCV_env.	
DR	InterPro; IPR002531; HCV_NS1.	
DR	InterPro; IPR000745; HCV_NS4a.	
DR	InterPro; IPR001490; HCV_NS4b.	
DR	InterPro; IPR002868; HCV_NS5a.	
DR	InterPro; IPR002166; HCV_RdRp.	
DR	InterPro; IPR001650; Helicase_C.	
DR	InterPro; IPR004109; Peptidase_S29.	
DR	InterPro; IPR009003; Pept_Ser_Cys.	
DR	InterPro; IPR002518; Pept_U39_HCV_NS2.	
DR	InterPro; IPR007095; RNA_pol_DS_PS.	
DR	InterPro; IPR007094; RNA_pol_PSVir.	





DR InterPro: IPR002521; HCV\_core.  
DR InterPro: IPR002519; HCV env.  
DR InterPro: IPR002531; HCV NS1.  
DR InterPro: IPR000745; HCV\_NS4a.  
DR InterPro: IPR001490; HCV\_NS4b.  
DR InterPro: IPR002868; HCV\_NS5a.  
DR InterPro: IPR002166; HCV\_RdRP.  
DR InterPro: IPR001650; Helicase C.  
DR InterPro: IPR004109; Peptidase 329.  
DR InterPro: IPR009003; Pept\_Ser\_Cys.  
DR InterPro: IPR002518; Pept\_U35\_HCV\_NS2.  
DR InterPro: IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro: IPR007094; RNA\_pol\_PSVir.  
DR Pfam: PF01543; HCV capsid; 1.  
DR Pfam: PF01542; HCV\_core; 1.  
DR Pfam: PF01539; HCV env; 1.  
DR Pfam: PF01560; HCV\_NS1; 1.  
DR Pfam: PF01538; HCV\_NS2; 1.  
DR Pfam: PF02907; HCV\_NS3; 1.  
DR Pfam: PF01006; HCV\_NS4a; 1.  
DR Pfam: PF01001; HCV\_NS4b; 1.  
DR Pfam: PF01506; HCV\_NS5a; 1.  
DR Pfam: PF00271; Helicase C; 1.  
DR Pfam: PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; DEXDG; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3011 AA; 32722 MW; 293F91430A0D4067 CRC64;  
  
Query Match 48.5%; Score 2868; DB 2; Length 3011;  
Best Local Similarity 51.7%; Pred. No. 3.4e-163;  
Matches 616; Conservative 41; Mismatches 105; Indels 430; Gaps 17;  
  
QY 204 TGAARATTSGLTSLSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNDSPVPVQSFQ 262  
DB 1165 SGGPLLCAGHAVLFRAAVCTRGVAKAVDFIPVENLGTTRSPVFTDNDSPVAVQSFQ 1224  
  
QY 263 VAHLHAPTSGKSTKVPAAVAAQGYKVLNPSVAATLGFAYMSKAHGIDPNIRTGVRT 322  
DB 1225 VAHLHAPTSGKSTKVPAAVAAQGYKVLNPSVAATLGFAYMSKAHGIDPNIRTGVRT 1284  
  
QY 323 ITTGSPITYTYGKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDOAETAGARLV 382  
DB 1285 ITTGSPITYTYGKFLADGCGSGAYDIIICDECHSTDATSIILGIGTVLDOAETAGARLV 1344  
  
QY 383 VLATATPPGCVTVPHNIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 442  
DB 1345 VLATATPPGCVTVSHNIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKCDELA 1404  
  
QY 443 AKLVALGINAVAYRGLDVSIPTSGDVVVVATDALTMTGVTGDFSDVIDCNTCTQTQVDF 502  
DB 1405 AKLVALGINAVAYRGLDVSIPTSGDVVVVSTDALMTGVTGDFSDVIDCNTCTQTQVDF 1464  
  
QY 503 SLDPFTTETITLPODAVSRTOGRGTPGKPGIYRFVAPGERPSGMFDSVLCCEYDAG 562  
DB 1465 SLDPFTTETITLPODAVSRTOGRGTPGKPGIYRFVAPGERPSGMFDSVLCCEYDAG 1524  
  
QY 563 CANYELTPAETTVRLRAYMTNPGLPVQDHLFEWEGVFTGLTHIDAHFLSQTKQSGENLP 622  
DB 1525 CANYELTPAETTVRLRAYMTNPGLPVQDHLFEWEGVFTGLTHIDAHFLSQTKQSGENFP 1584  
  
QY 623 YLVAYQATVCARAQAPPPSDQWKKLRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 682  
DB 1585 YLVAYQATVCARAQAPPPSDQWKKLRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 1644  
  
QY 683 IMTCMSADLEVVTS-----ACSGKPAIIPDREVLRYE 714  
DB 1645 IMTCMSADLEVVTSVWLVGVLAAALAAAYCLSTGCVVIVGKRVLSGKPAIIPDREVLRYE 1704  
  
QY 715 FDEMERCQHLPIYEQGMMLAEQFKQALGL-----SRGKPAIVDPKDEVLYQYDENEEC 770  
DB 1705 FDEMERCQHLPIYEQGMMLAEQFKQALGLLOATSRHAE-VITPAVQTNWQKL-EVFWA 1762

QY 771 SQAAPYIEQAQVIAHQFKEKVLGLDNDQVV-----VTPDKELLVE----- 811  
DB 1763 KHMWNFISGIQYLAG-----LSTLPGNPAIASLMAFTAATVTSPLTGTQTLLENILGGWV 1816  
  
QY 812 -----APDEMEBCASKAALIEGQRMAMLEKSKIQGLLG----- 845  
DB 1817 AAQLAAPGAATAFVGAGLAGAAIGSVGLGKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876  
  
QY 846 -----ILRRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVP 881  
DB 1877 EDLVNLLPALSPGALVGVVCAILURRHVGPGEQAVQMMNRLIAFASRGNHVSPTHYVP 1936  
  
QY 882 S----- 882  
DB 1937 ESDAAARVTAISSLTVTQLLRHLQHWISSECTTPCGSGWLRDIMDWICEVLSDFKTWLK 1996  
  
QY 883 ----- 882  
DB 1997 AKLMPOLPGIPFVSCQGYRGVWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPRTCRNMW 2056  
  
QY 883 ----- 882  
DB 2057 SGTFPINAVTTGPTCLPAPNYKFAWRSABEYVIRRVGDHYVSGMTTDLNLCPCQI 2116  
  
QY 883 -----RSRRFA-----QALPVWARPDY----- 899  
DB 2117 PSPEFFTELDGVRHLRFAPCKPELLREEVSVFRVGLHEYPVGSQPCPEPDVAULTSMLT 2176  
  
QY 900 -----NPP-----LVET---WKK- 909  
DB 2177 DPSHITAEAGRRLARGSPSSMASSASQLSAPSLKATCATANHDSPAELIEANLLWROE 2236  
  
QY 910 -----PDYEPVHVH-----RSSRRFAQALPVWARPDPVNP 940  
DB 2237 MGNITRVSESNKVVILDSFDPLVAEEDEREVSVPAILKSRFRARALPVWARPDPVNP 2296  
  
QY 941 LVETWKKPDYEPVHVHGRKTKRNTNRPDQVFPGGGQIVGRRGPPPIPKARR 992  
DB 2297 LVETWKKPDYEPVHVH-----GCPLPPPPSPVPPPRK 2329  
  
RESULT 7  
Q9PMW9 PRELIMINARY; PRT; 3015 AA.  
ID Q9PMW9  
AC Q9PMW9  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99420396; PubMed-10489358; DOI=10.1006/viro.1999.9889;  
RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;  
RT "Hepatitis C virus: an infectious molecular clone of a second major  
RT genotype (2a) and lack of viability of intertypic 1a and 2a  
RT chimeras.";  
RL Virology 262:250-263 (1999).  
DR EMBL: AF177039; AAF01181.1; -;  
DR EMBL: AF177037; AAF01179.1; -;  
DR PIR: PS0326; PS0326.  
DR PIR: PS0327; PS0327.  
DR PIR: PS0328; PS0328.  
DR HSSP: P27958; 1HEI.  
DR GO: GO:0016021; C:integral to membrane; IEA.  
DR GO: GO:0019028; C:viral capsid; IEA.  
DR GO: GO:0019031; C:viral envelope; IEA.  
DR GO: GO:0005524; F:ATP binding; IEA.  
DR GO: GO:0008026; F:ATP-dependent helicase activity; IEA.





AC Q6IX04; 2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Polyprotein (Fragment).  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OX Hepacivirus.  
RN NCBI\_TaxID=11103;  
[1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-HCV1a.  
RA Brann T.W., Kottilil S., Polis M., Imamichi T.;  
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY615798; AAT44836.1; -.  
DR HSSP; P27958; 1A1V.  
DR GO: 0016021; C: integral to membrane; IEA.  
DR GO: 0019038; C: viral capsid; IEA.  
DR GO: 0019031; C: viral envelope; IEA.  
DR GO: 0005524; F: ATP binding; IEA.  
DR GO: 0008026; F: ATP-dependent helicase activity; IEA.  
DR GO: 0016787; F: hydrolase activity; IEA.  
DR GO: 0003723; F: RNA binding; IEA.  
DR GO: 0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO: 0008236; F: serine-type peptidase activity; IEA.  
DR GO: 0005198; F: structural molecule activity; IEA.  
DR GO: 0006350; P: translation; IEA.  
DR GO: 0019079; P: viral genome replication; IEA.  
DR GO: 0019087; P: viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH N.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NS1.  
DR InterPro; IPR000745; HCV\_NS4b.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_psvir.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF01542; HCV\_core; 1.  
DR Pfam; PF01539; HCV\_env; 1.  
DR Pfam; PF01560; HCV\_NS1; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF01006; HCV\_NS4a; 1.  
DR Pfam; PF01003; HCV\_NS4b; 1.  
DR Pfam; PF01506; HCV\_NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRP; 1.  
DR SMART; SM00487; HXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW Hydrolase; Nonstructural protein; Polyprotein; Transmembrane.  
FT NON\_TER 2908  
SQ SEQUENCE 2908 AA; 315737 MW; BF5A4BC591498A4F CRC64;  
  
Query Match 48.5%; Score 2866; DB 2; Length 2908;  
Best Local Similarity 52.3%; Pred. No. 4.2e-163;  
Matches 610; Conservative 38; Mismatches 99; Indels 419; Gaps 17;  
  
QY 224 SQNIQLIVDFIPVENLETTMRSPPVTPQSFQVAHLHAPTSGKSTKVPAAAY 283  
DB TRGAKAVDFIPVENLETTMRSPPVTPQSFQVAHLHAPTSGKSTKVPAAAY 1245  
QY 284 AQGYKVLVLPNSVAATLGFAYMSKAHGIDPNRTGVRTITGSPITYSYGKFLADGGC 343

Db 1246 AQGYKVLVLPNSVAATLGFAYMSKAHGIDPNRTGVRTITGSPITYSYGKFLADGGC 1305  
QY 344 SGGAYDIIICDECHSTDATSIIGITGTVLDOAETAGARLVVLATATPPGCVTPHPNIEEV 403  
Db 1306 SGGAYDIIICDECHSTDATSIIGITGTVLDOAETAGARLVVLATATPPGCVTPHPNIEEV 1365  
QY 404 ALSTTGIPYGYKAIPLVETKGGHLLFCHSKKCDLAALVALGINAVAYYRGLDVS 463  
Db 1366 ALSTTGIPYGYKAIPLVETKGGHLLFCHSKKCDLAALVALGINAVAYYRGLDVS 1425  
QY 464 IPTSGDVVVATDALMTGYTGDPSVIDCNTCTVTQTVDFSLDPTFTTETITLPODAVSRT 523  
Db 1426 IPTSGDVVVATDALMTGYTGDPSVIDCNTCTVTQTVDFSLDPTFTTETITLPODAVSRT 1485  
QY 524 QRRGRTGRGPGIYRFVAPGERPSGMFSSVLCSCYDAGCAWYELTETVRLRAYNNT 583  
Db 1486 QRRGRTGRGPGIYRFVAPGERPSGMFSSVLCSCYDAGCAWYELTETVRLRAYNNT 1545  
QY 584 PGLPVCQDHLFEFEGVETGLTHDAHLSQTKSGENLPYLVAQATVCABAQAPPPSMD 643  
Db 1546 PGLPVCQDHLFEFEGVETGLTHDAHLSQTKSGENLPYLVAQATVCABAQAPPPSMD 1605  
QY 644 QMWKCLIRLKPDLHGPTLLYRLGAVONEITLTHPTVKYINTCMSADLEVVTSS 696  
Db 1606 QMWKCLIRLKPDLHGPTLLYRLGAVONEITLTHPTVKYINTCMSADLEVVTSS 1665  
QY 697 -----ACSGKPAIIPDREVLRYREFDEMECECSOHLPIYIQGMMLA 735  
Db 1666 VLAALAAVCLSTGCVIIGRIVLSGKPAVDPREVLRYREFDEMECECSOHLPIYIQGMMLA 1725  
QY 736 EQFKQKALGL-----SRGKPAIIPDREVLRYREFDEMECECSOHLPIYIQGMMLA 786  
Db 1726 EQFKQKALGLLQATASQAE-VIAPAVQTNMOKLEAFWAKHMMNFISGQYLAGLSTLPG- 1783  
QY 787 FKEKVLGLINDQVVVTP-----DKEILYE-----APDEMECECSKAALI 826  
Db 1784 -NPAIASLMAPTAATAVTSPLTTSQTLLENILGWSVAQAALPAATAFVAGLAGAAGS 1842  
QY 827 BEGQMAELKSKIQGLLG-----IL 847  
Db 1843 GLGKVLVDILAGYAGVAGALVAFKIMSGEPPTTDLVNLPLAILSPGALVGVVCAAIL 1902  
QY 848 RRHVGPGEAGVQMNRLIAFASRGHNHVSPTHYVPS----- 882  
Db 1903 RRHVGPGEAGVQMNRLIAFASRGHNHVSPTHYVPSDAAARVTAISSLTVTXLLRLHQ 1962  
QY 883 ----- 882  
Db 1963 WVSSECTTPCSGSLRDIWDWICEVLSDFKTLWAKLMPQLPGIPFVSCQRYGVRWGD 2022  
QY 883 ----- 882  
Db 2023 GIMHTRCHGAEITGHVKNGTMRIVGPKTCRNMSGTFPINAYTTGCTPLPAPNYTAL 2082  
QY 883 -----RSRRA----- 888  
Db 2083 WRVSAEYVEIRRVGDEHYVTGMTADNLKCPQVSPSEFFTELDGVLRLHFAPCKPLL 2142  
QY 889 -----QALPVWARPDPY-----NPP----- 902  
Db 2143 DEVSFRVGLHDYPVGSQPLCPCEPEPDVAULTSMLTDPSSHITAARRRRLARGSPPEASSS 2202  
QY 903 -----LVET---WKK-----PDYPPPVVH 918  
Db 2203 ASQLSAPSLRATCTTNHDSPDALIEANLLWRQMGNI TRVESKNVILDSFDPLVAE 2262  
QY 919 G-----RSSRRFAQALPVWARPDPYVPLVETWKKPDYPPVHGRKTKRNTNR 966  
Db 2263 EDEREVSVAABILKRRRFTPALPIWARPDPYVPLVETWKKPDYPPVHGRKTKRNTNR 2317  
QY 967 RPQDVKFGGQIVGRROPPIPKARR 992

Db 2318 PPQ-----SPVPVPPRK 2329

RESULT 10

O36609 PRELIMINARY; PRT; 3011 AA.

AC O36609; (T-EMBLrel. 05, Created)

DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)

DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

DE Polyprotein.

OS Hepatitis C virus strain H77.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus; Hepatitis C virus type 1; Hepatitis C virus type 1a.

OX NCBI\_TaxID=63746;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H77;

RX MEDLINE=97385173; PubMed=9238047; DOI=10.1073/pnas.94.16.8738;

RA Yanagi M., Purcell R.H., Emerson S.U., Bukh J.;

RT "Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee.";

RT Proc. Natl. Acad. Sci. U.S.A. 94:8738-8743 (1997).

RL ENBL; AF011752; AAB67037.1; -.

DR PIR; A44150; A44150.

DR PIR; P00804; P00804.

DR PIR; P03026; P03026.

DR PIR; P03027; P03027.

DR PIR; P03028; P03028.

DR HSSP; P26664; 1HEI.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.

DR GO; GO:0003723; F:RNA binding; IEA.

DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P:transcription; IEA.

DR GO; GO:0019079; P:viral genome replication; IEA.

DR GO; GO:0019087; P:viral transformation; IEA.

DR InterPro; IPR000345; CytC\_heme\_BS.

DR InterPro; IPR001410; DEAD/DEAH\_N.

DR InterPro; IPR011545; DEAD/DEAH\_N.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV NS1.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_S29.

DR InterPro; IPR009003; Pept\_Ser\_Cys.

DR InterPro; IPR002518; Rept\_U39\_HCV\_NS2.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV NS1; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRp; 1.

DR SMART; SM00487; DEXDc; 1.

DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.

KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.

SQ SEQUENCE 3011 AA; 327262 MW; 10D1C9702CA9B5DC CRC64;

Query Match 48.4%; Score 2861; DB 2; Length 3011;  
Best Local Similarity 51.6%; Pred. No. 8.9e-163;  
Matches 615; Conservative 40; Mismatches 107; Indels 430; Gaps 17;

QY 204 TGGAAATTTSGLTSLSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNSSPPVQSFQ 262  
DB 1165 SGGPLCLPAGHAVGLFRAAVCTRGKAKAVDFIPVENLGTMRSPVFTDNSSPPVQSFQ 1224  
QY 263 VAHLHAPTSGSKSTKVPAAAYAGQYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTCVRT 322  
DB 1225 VAHLHAPTSGSKSTKVPAAAYAGQYKVLVLPNSVAATLGFAYMSKAHGDIPNIRTCVRT 1284  
QY 323 ITTGSPITTYGKFLADGCGSGGAYDIIICDECHSTDATSIILGIGTVLDQASTAGARLV 382  
DB 1285 ITTGSPITTYGKFLADGCGSGGAYDIIICDECHSTDATSIILGIGTVLDQASTAGARLV 1344  
QY 383 VLATATPPGSGVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKDELA 442  
DB 1345 VLATATPPGSGVTPHPNIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKKDELA 1404  
QY 443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVIDCNTCVTQVDF 502  
DB 1405 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVIDCNTCVTQVDF 1464  
QY 503 SLDPFTTITITLPODAVSTQRRGTGRGKPGIYRFVAPGERPSGFMFSSVLCYCYDAG 562  
DB 1465 SLDPFTTITITLPODAVSTQRRGTGRGKPGIYRFVAPGERPSGFMFSSVLCYCYDAG 1524  
QY 563 CAWYELTPAETTVRLRAYMNTGCLPVCODHLEFWEVGFTGLTHIDAHFSLSQTKSGENLP 622  
DB 1525 CAWYELTPAETTVRLRAYMNTGCLPVCODHLEFWEVGFTGLTHIDAHFSLSQTKSGENFP 1584  
QY 623 YLVAYQATVCARAQAPPSQDMQMKCLIRLKTPLHGTPTLLYRLGAVQNEITLTHPVTKY 682  
DB 1585 YLVAYQATVCARAQAPPSQDMQMKCLIRLKTPLHGTPTLLYRLGAVQNEITLTHPVTKY 1644  
QY 683 IMTMSADLEVTST-----ACSGKPAIIPREVLRYE 714  
DB 1645 IMTMSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIVGRIVLSGKPAIIPREVLRYE 1704  
QY 715 FDEMECSQHLPIYIEQGMMLAEQFKOKLGL-----SRGKPAIVPDKVELYQYDMEEC 770  
DB 1705 FDEMECSQHLPIYIEQGMMLAEQFKOKLGLLQTSRHAEE-VITPAVQTNWKL-EVFWA 1762  
QY 771 SQAAPYIEQAQVIAHQFKEKVLGLIDNDQV-----VTPDKELIYE----- 811  
DB 1763 KHMWNFISGIQVLG-----LSTLPNPAIASLMAFTAATVTSPLTTGOTLLFNILGGWV 1816  
QY 812 -----AFDEMECSKAAALIEEQGRVAMLMKSLKIQGLLG----- 845  
DB 1817 AAQLAAPGAATAFVGAGLAGAAATGVLGKLVLDILAGYAGAGVAGALVAFKIMSGEVPST 1876  
QY 846 -----ILRRHVGPGEAGVQVMNRLIAFASRGNHVSPTHYVP 881  
DB 1877 EDLVNLLPAILSPGALVGVVCAAILRRHVGPGEAGVQVMNRLIAFASRGNHVSPTHYVP 1936  
QY 882 S----- 882  
DB 1937 ESDVAARVTAISSLTVTQLRLRHQWISSECTTPCGSGSWLRDWDWICEVLSDFKTLWK 1996  
QY 883 ----- 882  
DB 1997 AKLMPQLPGIPFVSCQRYGVWRGDGIMHTRCHCGAEITHGVKNGAMRIVGPRTCRNMW 2056  
QY 883 ----- 882  
DB 2057 SGTFLINAYTTGPTCLPAPNYKFAWRSABEYVEIRRVGDFHYISGMTTDLNKCPCQI 2116  
QY 883 -----RSRRFA-----QALPVWARPDY----- 899

```
Db 2117 PSPEFFTELDGRLRRFAPPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPDVAVLTSMLT 2176
Qy 900 -----NPP-----LVET---WKK- 909
Db 2177 DPHITABEAGRLRAGSPSPMASSASQSLAPSCLKATCTANHDSPDAELIEANLLNRQE 2236
Qy 910 -----PDYEPVVHG-----RSSRFAQALPVWARPDPNPP 940
Db 2237 MGGNITRVESKENVVILDSFDPLVAEDEREVSVPAILRKSRFAPALPVWARPDPNPP 2296
Qy 941 LVETWKLPDYEPPVHGKTKWNTRRPDQVKFGGQIVGRCGPPPKARR 992
Db 2297 LVETWKLPDYEPPVHG-----GCPLPPRSPPPVPPRK 2329

RESULT 11
Q9ELS8 PRELIMINARY; PRT; 3011 AA.
AC Q9ELS8
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP Desai S.M., Devare S., Yamaguchi J.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF290978; AAG02099.1; -
DR PIR; A44150; A44150.
DR PIR; Q0804; Q0804.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P26664; 1HEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV core.
DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5b.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA pol DS PS.
DR InterPro; IPR007094; RNA pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
```

```
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327108 MW; A6BECF5A3B3EE13F CRC64;

Query Match 48.2%; Score 2851; DB 2; Length 3011;
Best Local Similarity 51.5%; Pred. No. 3.5e-162;
Matches 611; Conservative 43; Mismatches 113; Indels 420; Gaps 17;

Qy 204 TGGAAARTTSLGTSLSFG-ASQNIQLIVDFIPVENLETTMRSVPVFTDNNSPPPVPPSQF 262
Db 1165 SGGPLLCAGHAVGLFRAAVCTRGAKAVDFIPVENLETTMRSVPVFTDNNSPPPVPPSQF 1224
Qy 263 VAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGDIPNIRTGVRT 322
Db 1225 VAHLHAPTSGSKSTKVPAAVAAQGYKVLVLPNSVAATLFGAYMSKAHGVDPNIRTGVRT 1284
Qy 323 ITTGSPTITYTYGKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDOAETAGARLV 382
Db 1285 ITTGSPTITYTYGKFLADGGCGGAYDIIICDECHSTDATSILGIGTVLDOAETAGARLV 1344
Qy 383 VLATATPPGSVTVPHPNIEEVALSTTGEIPYKAIPLEVIKGRHILFCHSKKCDLA 442
Db 1345 VLATATPPGSVTVPHPNIEEVALSTTGEIPYKAIPLEVIKGRHILFCHSKKCDLA 1404
Qy 443 AKLVALGINAVAYYRGDLVSIVPTSGDVVVVATDALMTGYTGDSDVDCNTCVTQTVDF 502
Db 1405 AKLVALGINAVAYYRGDLVSIVPASGDVVVSTDALMTGTFDPPDVIDCNTCVTQTVDF 1464
Qy 503 SLDPTFTIETITLPQDAVSRTOGRGRGPGIYRFVAPGERPSGMPDSSVLCBECYDAG 562
Db 1465 SLDPTFTIETITLPQDAVSRTOGRGRGPGIYRFVAPGERPSGMPDSSVLCBECYDAG 1524
Qy 563 CAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGLTHIDAHFLSOTKSGENLP 622
Db 1525 CAWYELTPAETTVRLRAYMNTPLPVCDHLEFEGVFTGLTHIDAHFLSOTKSGENFP 1584
Qy 623 YLVAYQATVCARAQAPPPSDQMKLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 682
Db 1585 YLVAYQATVCARAQAPPPSDQMKLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 1644
Qy 683 IMTCMSADLEVVTS-----ACSGKPAIIPDREVLVRE 714
Db 1645 IMTCMSANPEVVTSTWLVGGVLAALAAAYCLSTGCWIVGRIVLSGKPAIIPDREVLVQE 1704
Qy 715 FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKKEVLVQYD----- 765
Db 1705 FDEMECSQHLPIYEQGMMLAEQFKQKALGL-----SRGKPAIIPDKKEVLVQYD----- 1763
Qy 766 EMEECQAAPYIEQAQVIAHQFKEKVLGLIINDQVVVTP---DKEILYE----- 811
Db 1764 HMMNFISGTQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTLNLFILGGWVAQA 1821
Qy 812 -----AFDEMECASKAALIEEGORMAEMLSKIQGLLG----- 845
Db 1822 APGAATAFVAGLAGAAGTGSVGLGKVLVDILAGYAGVAGVALVAFKIMSGEVPSTEDLVN 1881
Qy 846 -----ILRRHVGPGEVQVMMNRLIAFASRGNHVSPTHYVPS---- 882
Db 1882 LLPAILSPGALVGVWCVAAILRRHVGPGEVQVMMNRLIAFASRGNHVSPTHYVPSDA 1941
Qy 883 ----- 882
Db 1942 ARVTAIISLNTVTLRLRLRHQWISGECTTPCSGSWLRDINDWICEVLSDFKTLKAKLMP 2001
```

```
Qy 883 ----- 882
Db 2002 QLPGIPFVSCQGVYGVWGDGIMHTRCHGAEITGHVKNGTMRIVGPRTCRNWWSGTFP 2061
Qy 883 ----- 882
Db 2062 INAYTGPCTPLPAPNYKFALMRVSAEYVEIRVGDHFYVSGMTTDLNLCPCQIPSPPEF 2121
Qy 883 ----- 882
Db 2122 FTELDGVLRHFAFPCKPLLEEVSFRVGLHEYPVGSQLECEPEPDVAVLTSMLTDPDSHI 2181
Qy 900 ----- 909
Db 2182 TABAAGRRLARGSPPPSSMASSASQLSAPLSKATCTTNHSDPAELIEANLLWRQMGNGNI 2241
Qy 910 ----- 945
Db 2242 TRVESENKVVILDSFDPLVAEEDEREVSVPAILRKSRFARALPFWARPDYNPLLIETW 2301
Qy 946 KPDYEPVPVHGRKTRNTNRNRPQDVKFPGGQIVGRRGPPIPKARR 992
Db 2302 KEPDYEPVH-----GCPLPPRSPVPPPRK 2329

RESULT 12
Q03463
ID Q03463 PRELIMINARY; PRT; 3011 AA.
AC Q03463
DT 01-NOV-1996 (TReMBUrel. 01, Created)
DT 01-NOV-1996 (TReMBUrel. 01, Last sequence update)
DT 01-MAR-2004 (TReMBUrel. 26, Last annotation update)
DE Polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,
RA Miyakawa Y., Mayumi M.;
RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117120; PubMed=1335573;
RA Okamoto H., Kanai N., Mishihiro S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus
RT isolate (HC-J1) with high homology to USA isolates.";
RL Nucleic Acids Res. 20:6410-6410 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91013116; PubMed=2107112;
RA Okamoto H., Okada S., Sugiyama Y., Yotsumoto S., Tanaka T.,
RA Yoshizawa H., Tada F., Miyakawa Y., Mayumi M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL Jpn. J. Exp. Med. 60:167-177 (1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=94174722; PubMed=7510436;
RA Mink M., Benichou S., Madaule P., Tiollais P., Prince A.,
RA Inchauspe G.;
RT "Characterization and mapping of a B-cell immunogenic domain in
RT hepatitis E virus E2 glycoprotein using a yeast peptide library.";
RL Virology 200:246-255 (1994).
RN [5]
RP SEQUENCE FROM N.A.
RA Okamoto H.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
DB EMBL; D10749; BAA01582.1; -.
```

```
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR PIR; S40770; S40770.
DR HSSP; P26664; IHEI.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002522; HCV capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept Ser Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
SQ SEQUENCE 3011 AA; 327114 MW; 97E9052C0250463B CRC64;

Query Match 48.2%; Score 2850; DB 2; Length 3011;
Best Local Similarity 51.2%; Pred. No. 4.1e-162;
Matches 608; Conservative 44; Mismatches 115; Indels 420; Gaps 18;

Qy 204 TGGAAARTTSGLTSLFSPG-ASQNIQLIVDFIPVENLETTMRSPVFTDNSSPPVQSFQ 262
Db 1165 SGGLPLCPAGHVVGIFRAAVCTRGVAKAVDFIPVESLETTMRSPVFTDNSSPPVQSFQ 1224
Qy 263 VAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLFGAYMYSKAHGIDPNIRTVRT 322
Db 1225 VAHLHAPTGSKSTKVPAAAYAAQGYKVLVLPNSVAATLFGAYMYSKAHGIDPNIRTVRT 1284
Qy 323 ITTGSPITYTYGKFLADGCGSGGAYDIIICDCHSTDATSLIGTIVLDQAEAGARLV 382
Db 1285 ITTGSPITYTYGKFLADGCGSGGAYDIIICDCHSTDATSLIGTIVLDQAEAGARLV 1344
Qy 383 VLATATPPGSIIVPHANIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA 442
Db 1345 VLATATPPGSIIVPHANIEEVALSTTGEIPFYGKAIPLEVIKGRHLIFCHSKKCKDELA 1404
Qy 443 AKLVALGINAVAYRGLDVSIVPTSGDVVVVATDALMTGYTGFDSVIDCNTCVTVDF 502
```



```
Db 1405 AKLVAGNVAAYRGLDVSIPSTGDDVVVATDALMTGYTGFDSIDNCVTQTQVDF 1464
Qy 503 SLDPFTIETITLQDQAVSRQGRGRGPGIYRFVAPGRPSGMFDSVLCCECYDAG 562
Db 1465 SLDPFTIETITLQDQAVSRQGRGRGPGIYRFVAPGRPSGMFDSVLCCECYDAG 1524
Qy 563 CAYELTAPAEITVRLAYMNTPGPLVPCQDHLFEWEGVFTGLTHIDAHFLSOTKSGENLP 622
Db 1525 CAYELTAPAEITVRLAYMNTPGPLVPCQDHLFEWEGVFTGLTHIDAHFLSOTKSGENFP 1584
Qy 623 YLVAYQATVCARAQAPPSNDMMKCLIRLKPETHGPTLLYRLGAGVNEITLHPVTKY 682
Db 1585 YLVAYQATVCARAQAPPSNDMMKCLIRLKPETHGPTLLYRLGAGVNEITLHPVTKY 1644
Qy 683 IMTCWSADLEVVVTS-----ACSGKPAIIPDREVLVRE 714
Db 1645 IMTCWSADLEVVVTSVWLVGGVLAALAAAYCLSTGCVVIVGRVILSGRPAIIPDREVLVRE 1704
Qy 715 FDEMECSQHLPIYIEQGMMLAEQFKQKALGL-----SRGKPAIIPDKEVLVQQYD----- 765
Db 1705 FDEMECSQHLPIYIEQGMMLAEQFKQKALGLLQTASRAE-VIAPTQVTNNQKLEAFWAK 1763
Qy 766 EMERCSQAPYIEQAQVIAHQFKVKLGLINDQVVVTP-----DKEILYE----- 811
Db 1764 HMNFISGIQYLAGLSTLPG--NPAIASLMAFTAAVTSPLTTSQTLNFIILGWNVAQA 1821
Qy 812 -----AFDEMECASKAALTEEGORMAEMLSKIQGLLG----- 845
Db 1822 APGAATAPVGLGAGAAVGSVGLGRVLVIDIAGVAGVAGALVAFKIMSGELPSTEDLVN 1881
Qy 846 -----ILRRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPS----- 882
Db 1882 LLPAILSPGALVGVCAAILRRHVGPGEVAVQWNNRLIAFASRGNHVSPTHYVPSDAA 1941
Qy 883 ----- 882
Db 1942 ARVTAISSLVTQTLLRLHOWLSSESTPCSGSWLRDIDWDICEVLSDFKTLKTLKMP 2001
Qy 883 ----- 882
Db 2002 HLPGIPFVSCQHYGKVGWRGDGIMHTRCHCGAEITGHVKNGTMRIVGPKTCRNMMWSGTFF 2061
Qy 883 ----- 882
Db 2062 INAYTGPCTPLPAPNYTFALWRVSAEYVEIRRVGDFFHYVTGMTDNLKCPQVPSPEF 2121
Qy 883 -----RSRPA-----QALPWARPDY----- 899
Db 2122 FTLDGVLRLHFRFAPCKPLLEEVSVFRVGLHDYPVGSQPCPEPEPDVAVLTSMLTDPHSI 2181
Qy 900 -----NPD-----LVET---WKK----- 909
Db 2182 TAAAGRLRAGSPSESSASSQSLSKATCTINHDSPDAELTEANLLWRQMGNI 2241
Qy 910 -----PDYEPVPHG-----RSSRRFAQALPVWARPDYNPPLVETW 945
Db 2242 TRVSENKVILDSFDPLVAEEDEREISVPAEILKSRFTQALPIWARPDYNPPLIETW 2301
Qy 946 KKPDYEPVPHGRTKRNTRNQVQKFPFGGQIVGRRGPIPKARR 992
Db 2302 KKNYPVPHGCPPLP-----PPQ-----SPVPVPRK 2329

RESULT 13
Q9DIT6
ID Q9DIT6 PRELIMINARY; PRT; 3011 AA.
AC Q9DIT6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Genomic RNA for polyprotein gene.
OS Hepatitis C virus.
```

```
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OX Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21014673; PubMed=11115058;
RA Kumar U., Tuthill T., Thomas H.C., Monjardino J.,
RT "Sequence, expression and reconstruction of an HCV genome from a
RL J. Viral Hepat. 7:459-465(2000).
DR EMBL; AJ278830; CAC03609.1; -.
DR PIR; PS0326; PS0326.
DR PIR; PS0327; PS0327.
DR PIR; PS0328; PS0328.
DR HSSP; P27958; 1A1V.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR GO; GO:0019087; P:viral transformation; IEA.
DR InterPro; IPR000345; CytC_heme_BS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR011542; Flavi_DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; Peptidase_S29.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR InterPro; IPR002518; Pept_U39_HCV_NS2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF07652; Flavi_DEAD; 1.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; Transmembrane.
FT CHAIN 810 1026 non-structural protein 2.
FT CHAIN 1027 1657 non-structural protein 3.
FT CHAIN 1658 1711 non-structural protein 4a.
FT CHAIN 2 191 core protein.
FT CHAIN 1712 1972 non-structural protein 4b.
FT CHAIN 1973 2420 non-structural protein 5a.
FT CHAIN 2421 3011 non-structural protein 5b.
FT CHAIN 192 383 envelop protein 1.
FT CHAIN 384 809 envelop protein 2.
SQ SEQUENCE 3011 AA; 327406 MW; 7B6264A74A5452D3 CRC64;
```

Query Match		48.0%; Score 2835; DB 2; Length 3011;
Best Local Similarity		50.7%; Pred. No. 3.3e-161;
Matches 612; Conservative 46; Mismatches 130; Indels 420; Gaps 18;		
QY	183	RTTSFGVSLFAPGAKQETHVTGGAARTTSGLTSLFSPG-ASQNTQLIVDRTPVENLET 241
DB	1144	RRDGRGSLSPRPISVYLGSSGGPDLCPAGHAGVIFRAAVCTRGVAKAVEFVVENLET 1203
QY	242	TMRSPVFTDNSSPPVQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLG 301
DB	1204	TMRSPVFTDNSSAPPVQSFQVAHLHAPTSGSKSTKVPAAAYAAQGYKVLVLPNSVAATLG 1263
QY	302	FGAYMSKAHGDINIRGTVRTITITGSPITVSTYTKFLADGCGSGGAYDIIICDECHSTDA 361
DB	1264	FGAYMSKAHGDINIRGTVRTITITGSPITVSTYTKFLADGCGSGGAYDIIICDECHSTDA 1323
QY	362	TSILGIGTGLDQAEAGARLVLATATPGSVTVPHNTEBEVALSTTGEPFYGKAIPLE 421
DB	1324	TSILGIGTGLDQAEAGARLVLATATPGSVTVPHNTEBEVALSTTGEPFYGKAIPLE 1383
QY	422	VIKGGRHLIFCHSKKCKDELAALVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 481
DB	1384	VIKGGRHLIFCHSKKCKDDLAALVALGINAVAYRGLDVSVIPTSGDVVVVATDALMTG 1443
QY	482	YTGDFOSVIDCNTCVTQTVDFSLDPTFTTITLTPQDAVSRTOGRGTGRGKPGIYRFVA 541
DB	1444	FTGDFOSVIDCNTCVTQTVDFSLDPTFTTITLTPQDAVSRTOGRGTGRGKPGIYRFVA 1503
QY	542	PGERPGRMDFSVLCBCYDAGCAWYELTPAETTVRLRAYNNTPLPVCQDHLFEWGVFT 601
DB	1504	PGERPGRMDFSAVLCBCYDAGCAWYELTPAETTVRLRAYNNTPLPVCQDHLFEWGVFT 1563
QY	602	GLTHIDAHFLSOTKQSGENLPYLVAQVATCARAQAPPPSWDQWMLKRLKPTLHGPTP 661
DB	1564	GLTHIDAHFLSOTKQSGENLPYLVAQVATCARAQAPPPSWDQWMLKRLKPTLHGPTP 1623
QY	662	LLYRLGAVQNEITLTPVTKYIMTCSADLEVTSS----- 696
DB	1624	LLYRLGAVQNEITLTPVTKYIMTCSADLEVTSTVWLVGVLAAALAAAYCLSTGCVVIV 1683
QY	697	---ACSGKPAIIDREVLREFDEMECSOHLPIEQGMWLAOFKOKALGL-----SRGG 749
DB	1684	GRIVLSGKPIIIDQEVLYREFDEMECSOHLPIEIPGMWLAOFKOKALALLQTSRQA 1743
QY	750	KPAIVDPKEVLYQOYD-----EMECSQAAPTEQAQVIAHQFKEKVLGLINDQVVVTP 804
DB	1744	E-VIAVAVQTNWORLEAFWAKHWNFISGIQYLAGLSTLPG--NPALASLMAETAATVSP 1800
QY	805	---DKELIYE-----AFDEMECASKAALIEEGORMAEMLSKIQGLL 844
DB	1801	LTTSTQLLNFILGGWAAQLAAPGAATAFVGAGLAGAAIGSVGLGVLDILAGYGAGVA 1860
QY	845	G-----ILRRHYGPGEGAVQWNNRLI 865
DB	1861	GALVAFKIMSGEVPSTEDMVNLLPAILSPGALVGVVVCATILRRHYGPGEGAVQWNNRLI 1920
QY	866	AFASRGNHVSPTHVPS----- 882
DB	1921	AFASRGNHVSPTHVPSDDAAARVAILSSLTVTQLRLHQWVSSEITPCSGSWLRDI 1980
QY	883	----- 882
DB	1981	WDWICEVLSDFTWLKAKLMPQLPGIPLVSCQRYGVWQDGMVHTRCHCGAEITGHVK 2040
QY	883	----- 882
DB	2041	NGTMRIVGPKTCRNMMSGTFPINAYTTGCTPLPAPNYTFALMRVSAEYVEIRRVGDFH 2100
QY	883	-----RSRRFA-----QAL 891
DB	2101	YVTGMTDNLRCQVQPSPEFTLDCGVLRRFAPPCKPLLRDEVSFRVLGHDPVGSQ 2160
QY	892	FVWARPDY-----NPP----- 902
RESULT 14		
POLG_HCVH		
ID	POLG_HCVH	STANDARD; PRT; 3011 AA.
AC	P27958;	
DT	01-AUG-1992 (Rel. 23, Created)	
DT	01-AUG-1992 (Rel. 23, Last sequence update)	
DT	25-OCT-2004 (Rel. 45, Last annotation update)	
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);	
DE	Envelope Glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2	
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)	
DE	(EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)	
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein	
DE	NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein	
DE	NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].	
OS	Hepatitis C virus (isolate H) (HCV).	
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
OC	Hepacivirus.	
OX	NCBI_TaxID=11108;	
RN	[1]_TaxID=11108;	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=92052256; PubMed=1658800;	
RA	Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,	
RA	Prince A.M.;	
RT	"Genomic structure of the human prototype strain H of hepatitis C	
RT	virus: comparison with American and Japanese isolates.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).	
RL	[2]	
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.	
RX	MEDLINE=97331322; PubMed=9187654;	
RA	Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;	
RT	"Structure of the hepatitis C virus RNA helicase domain.";	
RL	Nat. Struct. Biol. 4:463-467(1997).	
RN	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.	
RX	MEDLINE=98154321; PubMed=9493270; DOI=10.1016/S0969-2126(98)00010-0;	
RA	Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,	
RA	Murcko M.A., Lin C., Caron P.R.;	
RT	"Hepatitis C virus NS3 RNA helicase domain with a bound	
RT	oligonucleotide: the crystal structure provides insights into the mode	
RT	of unwinding.";	
RL	Structure 6:89-100(1998).	
CC	-1- FUNCTION: Protease NS2 is responsible for the cleavage of NS2-NS3.	
CC	-1- FUNCTION: Protease NS3 is responsible for the cleavage of NS3.	
CC	NS4A, NS4B, NS4B-NS5A and NS5A-NS5B.	
CC	-1- FUNCTION: NS4A forms a complex with NS3 and is essential for the	
CC	activation of NS3.	
CC	-1- FUNCTION: NS5A seems to have a transcriptional activatory role.	
CC	-1- FUNCTION: NS5B is a RNA-dependent RNA polymerase that plays an	
CC	essential role in the virus replication.	
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral	
CC	precursor polyprotein, commonly with Asp or Glu in the P6	
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.	
CC	-1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +	
CC	{RNA} (N).	
CC	-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a	
CC	lipoprotein envelope. The envelope consists of two proteins: E1	
CC	and E2. The nucleocapsid is a complex of protein C and mRNA.	

CC -1- PTM: The structural proteins C, E1 and E2 are produced by  
 CC proteolytic processing by the host signal peptidases.  
 CC -1- SIMILARITY: Contains 1 peptidase S29 domain.  
 CC -1- SIMILARITY: Contains 1 peptidase U39 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; M67463; AAA45534.1; -;  
 CC PIR; A36814; GNVVCH.

DR PIR; A18; X-ray; A/B=1017-1214, C/D=1676-1698.  
 DR PDB; 1A1V; X-ray; A=1192-1667.

DR PDB; 1A1V; X-ray; A=1192-1667.  
 DR PDB; 1HEI; X-ray; -;

DR MEROPS; S29.001; -;  
 DR MEROPS; U39.001; -;

DR TRANSFAC; T04155; -;  
 DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV\_core.

DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRP.

DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR009003; Pept\_Ser\_Cys.

DR InterPro; IPR002518; Pept\_U39\_HCV\_NS2.  
 DR InterPro; IPR004109; PeptIdase\_S29.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.

DR InterPro; IPR01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.

DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR SMART; SM00487; DEXDC; 1.

DR 3D-structure; ATP-binding; Coat protein; Core protein;  
 DR Envelope protein; Glycoprotein; Helicase; Hydrolase;

KW Nonstructural protein; Polyprotein; RNA-directed RNA polymerase;  
 KW Serine protease; Transferase; Transmembrane.

FT INIT\_MET 1 1  
 FT Removed from capsid protein C by the  
 FT cellular aminopeptidase.

FT CHAIN 1 191  
 FT Envelope glycoprotein E1.

FT CHAIN 192 383  
 FT Envelope glycoprotein E2.

FT CHAIN 384 746  
 FT Protein P7.

FT CHAIN 747 809  
 FT Nonstructural protein NS2.

FT CHAIN 810 1026  
 FT Protease/helicase NS3.

FT CHAIN 1027 1657  
 FT Nonstructural protein NS4A.

FT CHAIN 1658 1711  
 FT Nonstructural protein NS4B.

FT CHAIN 1712 1972  
 FT Nonstructural protein NS5A.

FT CHAIN 1973 2420  
 FT Nonstructural protein NS5B.

FT CHAIN 2421 3011  
 FT Potential.

FT TRANSMEM 347 369  
 FT Charge relay system (By similarity).

FT ACT\_SITE 1083 1083  
 FT Charge relay system (By similarity).

FT ACT\_SITE 1107 1107  
 FT Charge relay system (By similarity).

FT ACT\_SITE 1165 1165  
 FT ATP (Potential).

FT NP\_BIND 1230 1237  
 FT DECH box.

FT SITE 1316 1319  
 FT N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 196 196  
 FT N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 305 305 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 417 417 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 423 423 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 430 430 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 448 448 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 476 476 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 532 532 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 540 540 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 556 556 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 576 576 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 623 623 N-linked (GlcNAc. .) (Potential).  
 FT CARBOHYD 645 645 N-linked (GlcNAc. .) (Potential).  
 FT STRAND 1032 1035  
 FT HELIX 1039 1048  
 FT STRAND 1050 1050  
 FT STRAND 1059 1063  
 FT STRAND 1068 1074  
 FT TURN 1075 1076  
 FT STRAND 1077 1081  
 FT HELIX 1082 1085  
 FT TURN 1086 1087  
 FT STRAND 1090 1092  
 FT TURN 1093 1094  
 FT STRAND 1095 1097  
 FT STRAND 1101 1103  
 FT TURN 1104 1107  
 FT STRAND 1108 1112  
 FT TURN 1115 1116  
 FT STRAND 1120 1120  
 FT STRAND 1122 1122  
 FT STRAND 1129 1133  
 FT TURN 1135 1136  
 FT STRAND 1139 1144  
 FT TURN 1147 1148  
 FT STRAND 1149 1157  
 FT HELIX 1158 1161  
 FT TURN 1162 1163  
 FT STRAND 1165 1166  
 FT TURN 1172 1173  
 FT STRAND 1176 1186  
 FT TURN 1187 1188  
 FT STRAND 1189 1197  
 FT HELIX 1198 1205  
 FT STRAND 1680 1687  
 FT STRAND 1693 1694  
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 47.7%; Score 2821; DB 1; Length 3011;  
 Best Local Similarity 51.3%; Pred. No. 2.3e-160;  
 Matches 611; Conservative 41; Mismatches 110; Indels 430; Gaps 17;

Qy 204 TGGAAATTSGTSLFSPG-ASQNIQLIVDFIPVENLETTMRSVPFTDSSPPVPSFQ 262  
 Db 1165 SGGELLCPTGAVGLFRAAVCTRGVAKAVDFIPVENLETTMRSVPFTDSSPPVPSFQ 1224

Qy 263 VAHLHAPTSGSKTKVPAAYAAQYKVLVLPNSVAATLGFAYMSKAGIDPNRTGVRT 322  
 Db 1225 VAHLHAPTSGSKTKVPAAYAAQYKVLVLPNSVAATLGFAYMSKAGVDPNRTGVRT 1284

Qy 323 ITTGSPTITYTYGKFLADGGCGGAYDIIDICDECHSTDTATSIILGIVTLDQAETAGARLV 382  
 Db 1285 ITTGSPTITYTYGKFLADGGCGGAYDIIDICDECHSTDTATSIILGIVTLDQAETAGARLV 1344

Qy 383 VLATATPPGSVTVPHNIEEVALSTTGEIIPYKAIPLVIKGGHLLIFCHSKKCCDELA 442  
 Db 1345 VLATATPPGSVTVSHNIEEVALSTTGEIIPYKAIPLVIKGGHLLIFCHSKKCCDELA 1404

Qy 443 AKLVALGINAVAYRGDLDSVPTSGDVVVVATDALMTGYTGDSDSDCNTCVTQTVD 502  
 Db 1405 AKLVALGINAVAYRGDLDSVPTSGDVVVVATDALMTGYTGDSDSDCNTCVTQTVD 1464

Qy 503 SLDPFTTETITLPQDAVSRTRGRGRKPGYRFRVAPGERPSGMFSSVLCYDAG 562.

Db 1465 SLDPTFTIETTLPODAVSTQRRGRTGRKPGIYRFVAPGERPSGNSVLCYCYDAG 1524  
Qy 563 CAWVELTTPAETTVRLRAYMNTPLPVCDHLEFWEGVFTGLTHIDAHFELSOTKQSGENLP 622  
Db 1525 CAWVELTTPAETTVRLRAYMNTPLPVCDHLEFWEGVFTGLTHIDAHFELSOTKQSGENFP 1584  
Qy 623 YLVAYQATVCARAQAPPPSDQMWKCLIRLKPTLHGPTPLLYRLGAVQNEITLTHPVTKY 682  
Db 1585 YLVAYQATVCARAQAPPPSDQWKCLIRLKPTLHGPTPLLYRLGAVQNEVTLTHPVTKY 1644  
Qy 683 IMTCMSADLEVVTSS-----ACSGKPAIIPDREVLYRE 714  
Db 1645 IMTCMSADLEVVTSSVWLVGGVLAALAAAYCLSTGCVVIVGRIVLSGKPAIIPDREVLYQE 1704  
Qy 715 FDEMESCQHLPIYEOGMMLAEOKKALGL-----SRGKPAIVPDKVELYQYDSEMEEC 770  
Db 1705 FDEMESCQHLPIYEOGMMLAEOKKALGLLOTASRHAEE-VITPAVQTWQKL-EVFWA 1762  
Qy 771 SQAAPYIEQAQVIAHOFKEKVLGLINDQVV-----VTPDKELIYE----- 811  
Db 1763 KHMNFISGIQYLAGE-----LSTLPCNPALISMAFTAATVTSPLTTGQTLLENILGGMV 1816  
Qy 812 -----AFDEMEECASKAALIEEGORMAEMLSKTIQGLG----- 845  
Db 1817 AQLAAPGAATAFVGAGLAGAALDSVGLGKVLVDILAGYGAGVAGALVAFKIMSGEVPST 1876  
Qy 846 -----ILRRHVGPGEVAVQNMNRLIAFASRGNHVSPTHYVP 881  
Db 1877 EDLVNLLPALSPGALAVGVFASILRRRVGPGEVAVQNMNRLIAFASRGNHVSPTHYVP 1936  
Qy 882 S----- 882  
Db 1937 ESDAARVTAISSLVTQLLRHLQHWISSECTPCSGSLRDIWDWICEVLSDFKTLWK 1996  
Qy 883 ----- 882  
Db 1997 AKLMPQLPGIPFVSCQYRGVNRGDIIMHTRCHGAEITHGVKNGTMRIVGPRCTCKNMW 2056  
Qy 883 ----- 882  
Db 2057 SGTFFINAYTGTCTELPAPNYKALWRVSAEYVEIRRVGDFHYVSGMTDNLKPCQI 2116  
Qy 883 -----RSRFA-----QALPVWARPDY----- 899  
Db 2117 PSPEFTELDGVLRHFRAPCKPLLRREVSFRVGLHEYPVGSQLPCEPEPDVAVLSMLT 2176  
Qy 900 -----NPP-----LVET---WKK- 909  
Db 2177 DPSHITAAGRRRLARGSPPSMASSASQLSAPSLKATCTANHDSPPDAELIEANLLWRQE 2236  
Qy 910 -----PDYEPVWHG-----RSSRFAQALPVWARPDPNPP 940  
Db 2237 MGNITRVSEKNKVLDSFDPLVAEEDEREVSVPAILRKSRFAPALPVWARPDPNPL 2296  
Qy 941 LVETWKKPDYEPVWHGRKTKRNTNRPPQDVKFPFGGQIVGRRGPPPIPKARR 992  
Db 2297 LVETWKKPDYEPVWH-----GCPLPPSPPPVPPRK 2329

RESULT 15  
Q9J3H0  
ID Q9J3H0 PRELIMINARY; PRT; 3010 AA.  
AC Q9J3H0;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polyprotein.  
OS Hepatitis C virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepacivirus.  
OX NCBI\_TaxID=11103;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=MD22;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF207763; AAF65953.1; -;  
DR PIR; A61196; A61196.  
DR PIR; PQ0246; PQ0246.  
DR PIR; PQ0254; PQ0254.  
DR PIR; PS0329; PS0329.  
DR HSSP; Q8JYSL; 1CWX.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0019028; C:viral capsid; IEA.  
DR GO; GO:0019031; C:viral envelope; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006350; P:transcription; IEA.  
DR GO; GO:0019079; P:viral genome replication; IEA.  
DR GO; GO:0019087; P:viral transformation; IEA.  
DR InterPro; IPR000345; CytC\_heme\_BS.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR011545; DEAD/DEAH N.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002531; HCV NS1.  
DR InterPro; IPR000745; HCV NS4a.  
DR InterPro; IPR001490; HCV NS4b.  
DR InterPro; IPR002868; HCV NS5a.  
DR InterPro; IPR002166; HCV RdRp.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_S29.  
DR InterPro; IPR009003; Pept\_Ser\_Cys.  
DR InterPro; IPR002518; Pept\_U39 HCV NS2.  
DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_pol\_PSVir.  
DR Pfam; PF01543; HCV capsid; 1.  
DR Pfam; PF01542; HCV core; 1.  
DR Pfam; PF01539; HCV env; 1.  
DR Pfam; PF01560; HCV NS1; 1.  
DR Pfam; PF01538; HCV NS2; 1.  
DR Pfam; PF02907; HCV NS3; 1.  
DR Pfam; PF01006; HCV NS4a; 1.  
DR Pfam; PF01001; HCV NS4b; 1.  
DR Pfam; PF01506; HCV NS5a; 1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00998; Viral\_RdRp; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
KW Polyprotein; Transmembrane.  
SQ SEQUENCE 3010 AA; 32709 MW; 737BEF31E3C2B28D CRC64;  
Query Match 47.1%; Score 2783.5; DB 2; Length 3010;  
Best Local Similarity 48.3%; Pred. No. 4.1e-158;  
Matches 590; Conservative 59; Mismatches 125; Indels 448; Gaps 14;  
Qy 183 RTTSGFVSLFAPGAKQKNETHTVTGGAARTTSGLTSLFSPG-ASQNTQLIVDFIPVENLET 241  
Db 1144 RRGDGRGSLSPRPVSYLKGSSGGPLLCPLGHVVGIFRAAVCTRGVAKAVDFVPVESMET 1203  
Qy 242 TMRSPVFTNSSPVPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGVKLVLPNSVAATLG 301  
Db 1204 TMRSPVFTNSSPVPVQSFQVAHLHAPTGSGKSTKVPAAAYAAQGVKLVLPNSVAATLG 1263  
Qy 302 FGAYMSKAGIDPNIRTNRTTGTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDA 361  
Db 1264 FGAYMSKAGIDPNIRTNRTTGTGSPITYSTYTGKFLADGCGSGGAYDIIICDECHSTDS 1323



1000